

Application of image processing and knowledge reasoning in the construction of new power system

Edited by

Fuqi Ma, Bo Wang, Hengrui Ma, Yikui Liu and Rong Jia

Published in

Frontiers in Energy Research



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-6088-4
DOI 10.3389/978-2-8325-6088-4

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Application of image processing and knowledge reasoning in the construction of new power system

Topic editors

Fuqi Ma — Xi'an University of Technology, China

Bo Wang — Wuhan University, China

Hengrui Ma — Qinghai University, China

Yikui Liu — Sichuan University, China

Rong Jia — Xi'an University of Technology, China

Citation

Ma, F., Wang, B., Ma, H., Liu, Y., Jia, R., eds. (2025). *Application of image processing and knowledge reasoning in the construction of new power system*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6088-4

Table of contents

05	Transmission line bolts and their defects detection method based on position relationship Zhenbing Zhao, Jing Xiong, Yu Han and Siyu Miao
14	A two-stage distributed optimization method for home energy management systems via multi-modal data-driven algorithm Feifei Cui, Dou An and Yingzhuo Zhao
31	Defect detection method for key area guided transmission line components based on knowledge distillation Zhenbing Zhao, Xuechun Lv, Yue Xi and Siyu Miao
41	The application of a lightweight model FA-YOLOv5 with fused attention mechanism in insulator defect detection Jun Liu, Mingming Hu, Junyuan Dong and Xu Lu
56	Short-term interval prediction of PV power based on quantile regression-stacking model and tree-structured parzen estimator optimization algorithm Hongyang Zhang, Rong Jia, Haodong Du, Yan Liang and Jiangfeng Li
70	An efficient YOLO v3-based method for the detection of transmission line defects Changbao Xu, Mingyong Xin, Yu Wang and Jipu Gao
81	Search strategy and line association analysis of cascading failure accident chain in new energy power systems Xueting Cheng, Wenxu Liu, Yueshuang Bao and Xinyuan Liu
93	Smart grid power load type forecasting: research on optimization methods of deep learning models Huadong Sun, Yonghao Ren, Shanshan Wang, Bing Zhao and Rui Yin
110	Optimal scheduling of microgrids considering real power losses of grid-connected microgrid systems Shengbo Sun, Wei Guo, Qiwei Wang, Peng Tao, Gang Li and Zhenbing Zhao
119	Research on large-scale clean energy optimal scheduling method based on multi-source data-driven Chuanyu Xiong, Lingfeng Xu, Li Ma, Pan Hu, Ziyong Ye and Jialun Sun
128	Decentralized asynchronous adaptive federated learning algorithm for securely prediction of distributed power data Qiang Li, Di Liu, Hui Cao, Xiao Liao, Xuanda Lai and Wei Cui
142	A comprehensive evaluation method for moisture and aging state of oil-immersed paper bushings based on dielectric modulus and K-nearest neighbor algorithm Zhengping Liang, Yongbin Sun, Hao Cheng, Na Zhang, Bo Li, Yang Liu, Yan Fang and Yiyi Zhang

- 158 **A novel recognition method for complex power quality disturbances based on Markov transition field and improved densely connected network**
Lei Zhou, Shuifu Gu, Yi Liu and Chaoqun Zhu
- 174 **Knowledge reasoning in power grid infrastructure projects based on deep multi-view graph convolutional network**
Jie Hu, Gang Xu, Lizhong Qi and Xin Qie
- 185 **Research on detection of transmission line corridor external force object containing random feature targets**
Hongbo Zou, Ziyong Ye, Jialun Sun, Junting Chen, Qinhe Yang and Yanhui Chai
- 197 **A transformer acoustic signal analysis method based on matrix pencil and hybrid deep neural network**
Qizhe Zhang, Guozheng Peng, Yuanpeng Tan, Zhonghao Zhang and Xiaojing Bai
- 207 **A bolt defect detection method for transmission lines based on improved YOLOv5**
Hongbo Zou, Jialun Sun, Ziyong Ye, Jinlong Yang, Changhua Yang, Fengyang Li and Li Xiong
- 219 **Abnormal line loss identification and category classification of distribution networks based on semi-supervised learning and hierarchical classification**
Wei Li, Wen Zhao, Junmin Li, Jie Li and Yankai Zhao
- 233 **Terminal strip detection and recognition based on improved YOLOv7-tiny and MAH-CRNN+CTC models**
Guo Zhijun, Luo Weiming, Chen Qiujie and Zou Hongbo
- 244 **A fusion topology method for generating new equipment startup schemes for power grids**
Tao Meng, Xiaohui Lu, Xiaoang Wang, Liang Wang, Wenbiao Tao, Lianfei Shan and Xiaofei Geng
- 256 **A causal reasoning approach for power transformer failure diagnosis**
Fei Jiao, Zhenyuan Ma, Qikun Chen, Fengda Zhang and Dezong Zhao
- 266 **LoadNet: enhancing energy storage system integration in power system operation using temporal convolutional and recurrent models with self-attention**
Minggang Liu and Xiaoxu Hu
- 283 **Drone image recognition and intelligent power distribution network equipment fault detection based on the transformer model and transfer learning**
Jiayong Zhong, Yongtao Chen, Jin Gao and Xiaohong Lv



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Shuaiqi Liu,
Hebei University, China
Tadachika Ozono,
Nagoya Institute of Technology, Japan
Xian Tao,
Chinese Academy of Sciences (CAS),
China

*CORRESPONDENCE

Siyu Miao,
✉ siyumiao_hbu@163.com

RECEIVED 29 July 2023

ACCEPTED 05 September 2023

PUBLISHED 14 September 2023

CITATION

Zhao Z, Xiong J, Han Y and Miao S (2023),
Transmission line bolts and their defects
detection method based on
position relationship.
Front. Energy Res. 11:1269087.
doi: 10.3389/fenrg.2023.1269087

COPYRIGHT

© 2023 Zhao, Xiong, Han and Miao. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Transmission line bolts and their defects detection method based on position relationship

Zhenbing Zhao^{1,2,3}, Jing Xiong^{1,4}, Yu Han¹ and Siyu Miao^{1*}

¹School of Electrical and Electronic Engineering, North China Electric Power University, Baoding, China, ²Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education, North China Electric Power University, Baoding, China, ³Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding, China, ⁴Department of Information Engineering, Sichuan Vocational and Technical College of Communications, Chengdu, China

Introduction: To solve the problems of small proportion of bolts in aerial images of power transmission lines, small differences between classes, and difficulty in extracting refined features, this paper proposes a method for detecting power transmission line bolts and their defects based on positional relationships.

Methods: Firstly, a spatial attention module is added to Faster R-CNN, using two parallel cross attention to obtain cross path features and global features respectively, and spatial feature enhancement is performed on the features output from the convolution layer. Then, starting from the spatial position relationship of bolts and their defects, using the relative geometric features of candidate regions as input, the spatial position relationship of bolts and their defects on the image is modeled. Finally, the position features and regional features are connected to obtain enhanced features. The bolt position knowledge on the connecting plate is added to the detection model to improve the detection accuracy of the model.

Results and discussion: The experimental results show that the mAP value of the algorithm in this paper is increased by 6.61% compared to the Faster R-CNN detection model in aerial photography of transmission line bolts and their defect datasets, with the AP value of normal bolts increased by 1.73%, the AP value of pin losing increased by 4.45%, and the AP value of nut losing increased by 13.63%.

KEYWORDS

transmission line bolts, bolts defects, target detection, attention mechanism, positional relationship

1 Introduction

Under the urgent requirement of achieving carbon peaking and carbon neutrality goals, the structure of China's power system form is undergoing fundamental changes (Zhuo et al., 2023). With the construction and development of new power systems, new power equipment such as power electronics and large-scale energy storage devices will be widely used, and the requirements for flexible and controllable and safe and stable power grids are becoming increasingly high (Sheng et al., 2021). Bolts are present in large numbers in power transmission lines and play a key role in fixing and connecting various components. However, due to long-term working in the wild, various components will be affected by natural environment and the external mechanical load tension and the internal power load of the power system. These factors can cause the bolts connected to various components to

produce defects such as loose nuts, losing, and stripped pins, which seriously affects the stable transmission of electrical energy (Zhao et al., 2021; Zhao and Ding, 2022). Therefore, in order to ensure the safe operation of transmission lines, it is crucial to carry out regular inspection of transmission line bolts and their defects.

Traditional transmission line inspection methods require electric workers to climb the tower, and in order to accurately record the cause of faults and carry out timely repairs, workers need to carry a variety of working tools when climbing the tower, which is not only economically costly, dangerous working environment and extremely inefficient inspection. Therefore, with the continuous development of deep learning technology, more and more scholars have started to adopt deep learning for the analysis and processing of transmission line images. Artificial intelligence technology has become a current hotspot and has performed powerfully in many fields, especially in target detection with many excellent results (Ge et al., 2017). And the main task of target detection is to find out all the targets of interest in the image and determine their class and location. Deep learning-based target detection algorithms are mainly divided into two categories, one is the Faster R-CNN (Ren et al., 2017) (Faster Regions with Convolutional Neural Network Features) series, a two-stage detection model based on region suggestion. One is a regression-based single-stage detection model based on the

YOLO (Redmon et al., 2016) (You Only Look Once) series and the SSD (Liu et al., 2016) (Single Shot Detection) series.

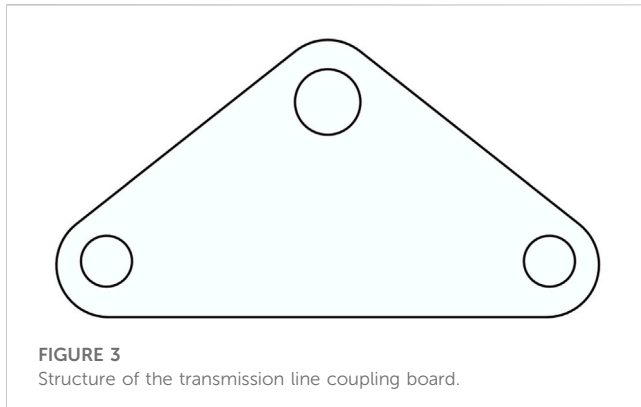
To achieve automatic detection of transmission line bolts and their defects, many scholars have introduced target detection algorithms into the detection of bolts and their defects to reduce the reliance on traditional manual inspection. A transmission line bolt detection method for processing massive UAV (Unmanned Aerial Vehicle) image data using UAV inspection images was proposed by Feng et al. (2018). This method firstly establishes a sample library, extracts HOG (Histogram of Oriented Gradients) features and constructs SVM (Support Vector Machine) classification to achieve recognition of high-resolution UAV inspection images. However, this method is susceptible to the influence of image illumination and effects. A deep learning-based transmission line bolt detection system for transmission line bolts with inconspicuous features, small size and difficult detection in inspection images was proposed by Zhang et al. (2021). This method adopts the principle of hierarchical detection, using the SSD algorithm to locate the defective bolt connection parts and cut out the connection parts, increasing the proportion of bolts in inspection images. Secondly, this method uses data augmentation to expand the dataset, and finally uses the YOLOv3 algorithm to detect defective bolts. An automatic detection model called Automatic Visual Shape Clustering Network (AVSCNet) was constructed to detect losing pins by Zhao et al. (2020). First, an unsupervised clustering method for bolometric visual shapes is proposed and applied to construct a defect detection model that learns differences in visual shapes. Next, three deep convolutional neural network optimisation methods are used in the model: feature enhancement, feature fusion, and region feature extraction. Regression calculation and classification are applied to the region features to obtain defect detection results. However, during the training of the model, many hyperparameters need to be set manually and do not have automatic learning capabilities. To solve the problems of too small bolt targets, small differences between different categories and difficulty in extracting fine features, a detection method with a dual attention mechanism was proposed by Qi et al. (2021). This method analyzes and enhances visual features at different scales and locations respectively. This method uses multi-scale attention modules to enhance fine features in the bolt region and spatial attention modules to increase the feature differences



FIGURE 1
Aerial photograph of the original bolts of the transmission line.



FIGURE 2
Transmission line bolts and their defects.



between the bolt and the background to improve the prediction of the bolt region. The Ultrasmall Bolt Defect Detection Model (UBDDM) based on Deep Convolutional Neural Network (DCNN) was proposed by Luo et al. (2023). Which included Ultrasmall Object Perception Module (UOPM) and Local Bolt Detection Module (LBDM), and introduces a hybrid attention mechanism and multi-scale feature fusion to further improve the network's ability to extract shallow features. A novel and high-accuracy defect detection method based on deep learning technology, named insulator defect detection network (I2D-Net) was proposed by Fu et al. (2023), which improves the ability of defect location in the presence of interference factors. An efficient and high-performance defect detection model called DDNet is proposed by Gong et al. (2023) to recognize defects from images of unmanned aerial vehicles. The attention mechanism was adopted in the improved detection model in order to enhance the representation learning of the image. However, the model only focuses on

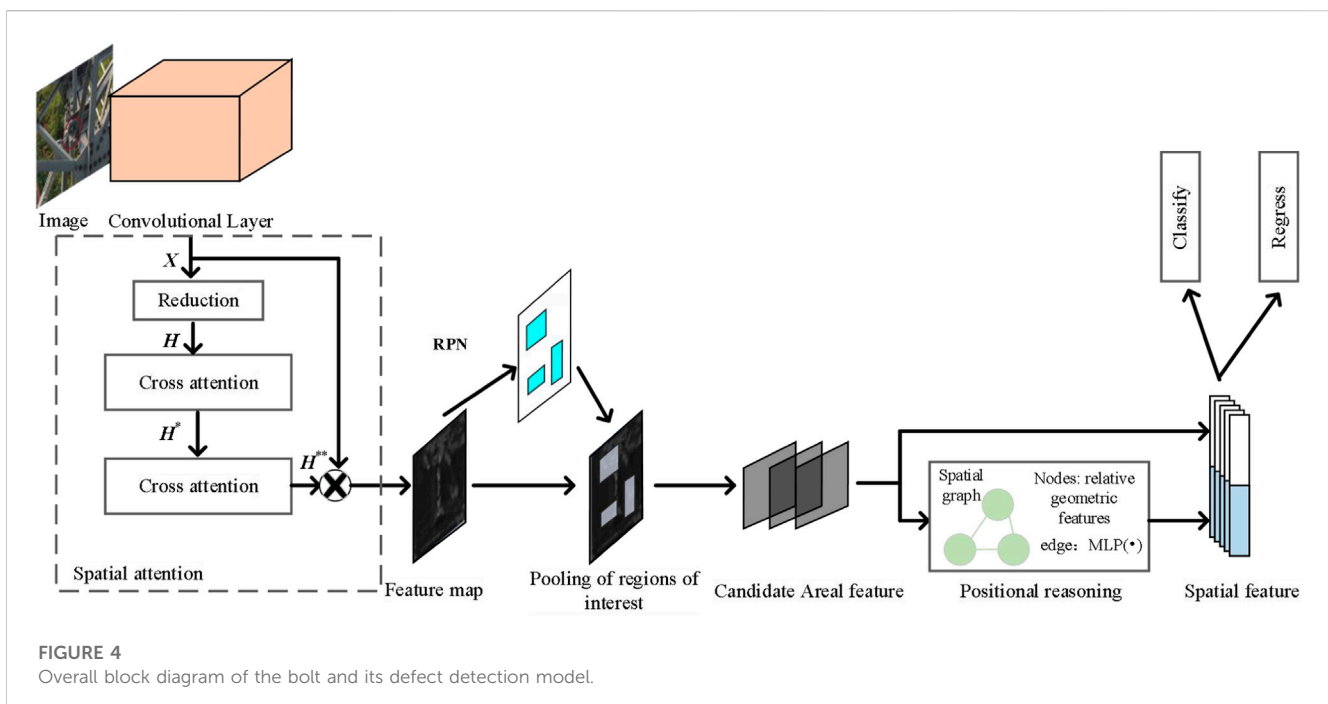
enhancing the bolt features, without incorporating the inherent bolt position information into the detection.

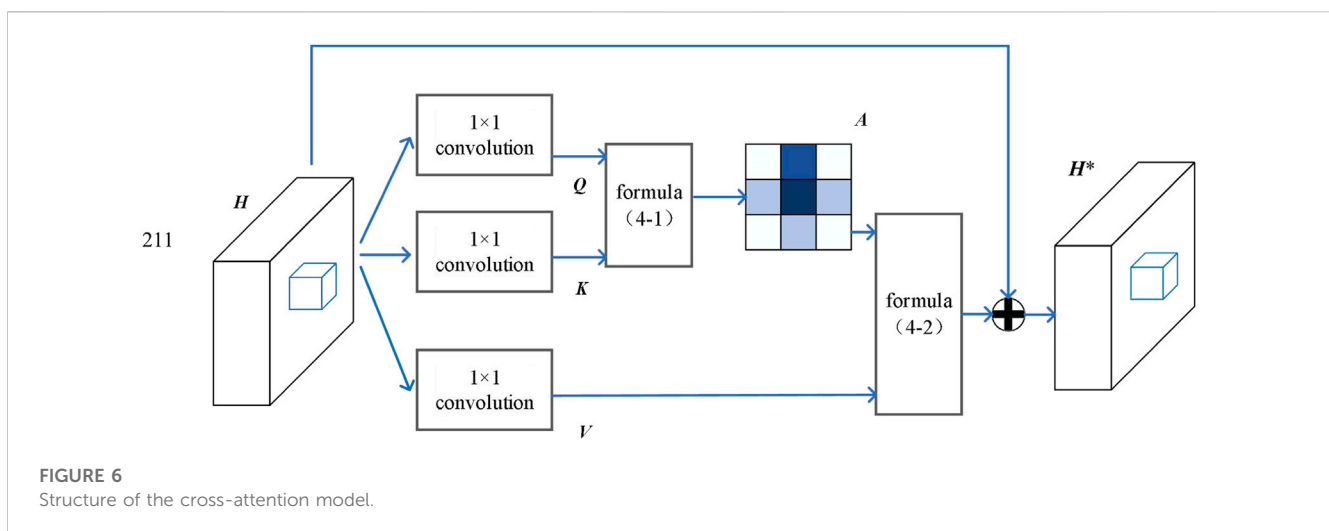
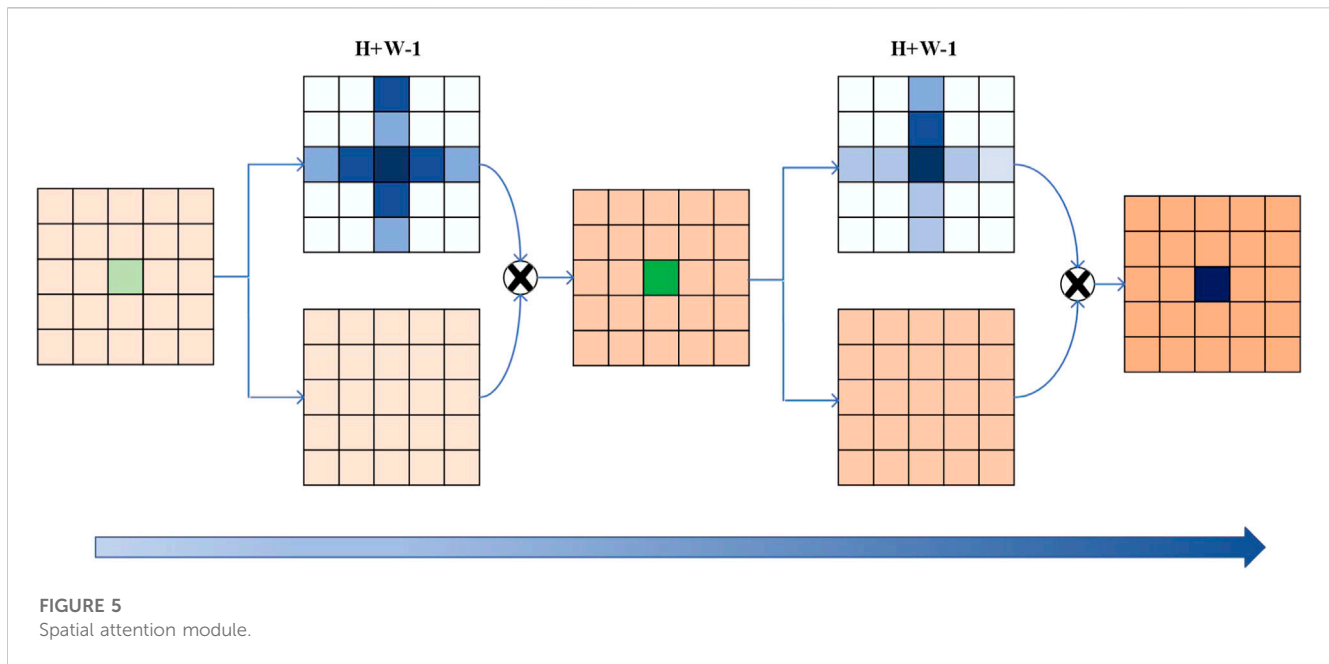
Transmission lines have problems such as small targets for bolts and their defects, limited information on targets, difficulty in feature extraction, and small differences between target classes. To alleviate these problems to a certain extent and promote the widespread application of deep learning in the detection of bolts and their defects in transmission lines, this article proposes the following methods. Firstly, the spatial attention module is added to the Faster R-CNN model to help the model acquire global features to achieve the purpose of bolts and their defects feature enhancement. Then the location inference module is used to add location relations to the detection model to reduce the leakage and false detection of bolts and their defects detection, which can improve the detection effect of the model.

2 Research background

As a crucial connection component, bolts are present in large numbers on all types of fixtures, which are widely distributed and numerous in transmission lines. However, the bolts and their defects account for a very small proportion of the aerial images, making them easy to miss detection when they are directly detected together with the fixtures. In Figure 1, the original bolt image is nearly impossible to identify within the complicated background of the aerial image. The detection model is often at risk of losing important features when extracting bolt features, rendering bolt detection a challenging task.

Therefore, this paper discards the idea of directly detecting bolts and their defects on aerial images, but instead annotates the metal tool targets, mainly joint plates, to build a dataset of bolts and their defects. Afterwards, carries out the detection of transmission line bolts and their defects afterwards. This design method can effectively



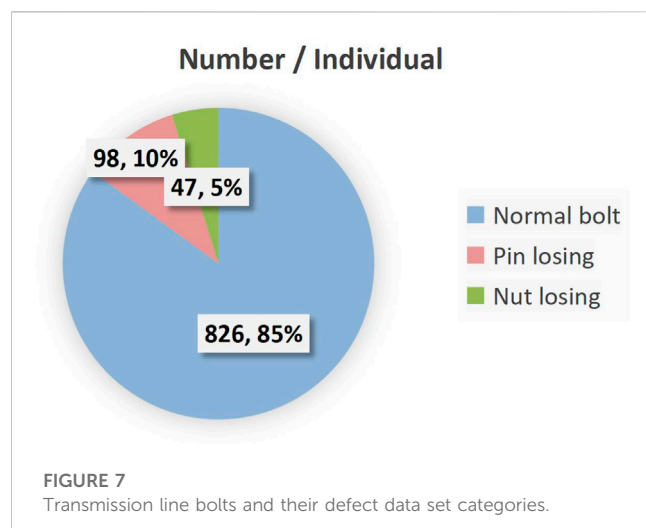


increase the proportion of bolt targets in the image, reduce the influence of complex background on bolt detection, and effectively reduce the information loss of bolt targets. We add a spatial attention module to the detection model to improve the model's ability to extract bolts and their defect features.

Figure 2 shows several sets of cropped yoke plate images. It can be seen that the bolts and their defect targets in the image are distributed in a triangular pattern, and the cropped image is clearer than the original aerial image. The cropped image size is smaller, which can make the detection speed faster.

As the bolts on the transmission line coupling plate generally have a fixed position, the coupling plate target is chosen for the labeling of transmission line bolts and their defect data sets. Figure 3 shows the structure of the transmission line joint board, which is a board-shaped

connection for the parallel assembly of multiple branches, mostly used for the parallel assembly of double insulator strings and multiple insulator strings, the assembly of insulator strings with double and multiple wires and the assembly of double pulling wires and other connections. There are various types of plates, such as L-plate, LZ-plate, LF-plate, LJ-plate and LE-plate, etc. The different types of plates are subject to different forces depending on their structure. In this paper we primarily focus on L-plates. The coupling plate in Figure 3 is the most typical and common L-shaped coupling plate, which is used for assembling double tension insulator strings with a single conductor, single insulator strings with two split conductors, and also for forming triple insulator strings in parallel. In Figures 3A–C are bolt installation positions. It can be found that the position of the bolts on the L-shaped coupling plate is fixed and the connection of the



midpoints of each bolt can form a triangle. Therefore, *a priori* knowledge can be added to the detection of the bolts and their defects, and the inference of the bolt position relationship can be added after the (Region Proposal Network) RPN generates the candidate region in order to improve the fit of the bolt detection frame to the target, and at the same time improve the detection accuracy of the model.

3 Materials and methods

The overall block diagram of the bolt and its defect detection model in this paper is shown in Figure 4. Firstly, to address the problem of small differences between classes of bolts and their defects and small bolt targets, spatial attention is added after the feature extraction network ResNet-101 to help capture global dependencies for each pixel through two crossover networks, so that the bolts and their defect targets can obtain global contextual information efficiently and quickly. The specific process is divided into two branches. One is to obtain the feature map H by convolutional downscaling of the features output from the convolutional layer, and then input H into the cross-attention module to produce a new feature map H^* . At this point, H^* contains horizontal and vertical contextual information, after which H^* is inputted again into the cross-attention module to output the feature map H^{**} . The other branch is to keep the output features unchanged. Afterwards, the features of the two branches are fused together to obtain global contextual information. Secondly, to address the lack of inference capability for bolt and its defect detection, a location relationship inference module is added after the model generates the box of interest. The main approach is to take the bounding box of the candidate region as input, learn the spatial discrepancy of the region as the edge of the region node, then output the location relationship knowledge to enhance the location features, and finally connect the location features with the region features for regression classification process to obtain the final detection results.

TABLE 1 Ablation experiment.

Methods	Normal bolt AP (%)	Pin losing AP (%)	Nut losing AP (%)	mAP (%)
Baseline	88.95	60.86	33.74	61.18
Baseline + spatial attention	89.57	54.90	46.57	63.68
Baseline + positional reasoning	89.40	65.21	43.43	66.02
Ours	90.68	65.31	47.37	67.79

TABLE 2 Model detection results before and after improvement of the backbone network for VGG16.

Methods	Normal bolt AP (%)	Pin losing AP (%)	Nut losing AP (%)	mAP
Faster R-CNN	88.62	28.75	26.26	47.88
Ours	88.06	37.37	74.71	66.71

TABLE 3 Multiple model detection accuracy.

Methods	Normal bolt AP (%)	Pin losing AP (%)	Nut losing AP (%)	mAP (%)
RetinaNet	95.7	23.2	56.4	58.4
Faster R-CNN	89.0	60.9	33.7	61.2
Cascade R-CNN	89.9	77.9	27.3	65.0
Sparse R-CNN	87.1	54.2	60.6	67.3
Ours	90.7	65.3	47.4	67.8

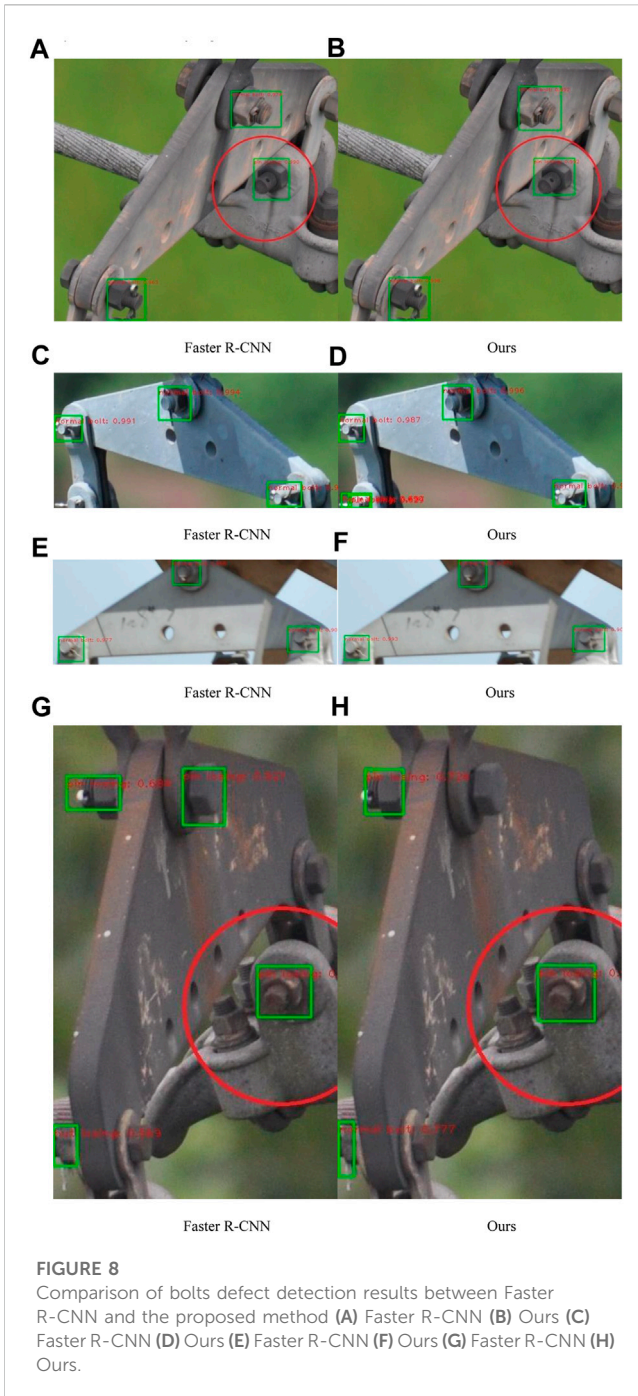


FIGURE 8

Comparison of bolts defect detection results between Faster R-CNN and the proposed method (A) Faster R-CNN (B) Ours (C) Faster R-CNN (D) Ours (E) Faster R-CNN (F) Ours (G) Faster R-CNN (H) Ours.

3.1 Spatial attention

Bolts as a large number of fixed connection components in the transmission line, their distribution exists in a certain pattern, especially in the bolts on specific fixtures, the location of the bolts is more basically fixed. The original Faster R-CNN only focuses on the local area of the image when detecting, without considering the spatial pattern of the transmission line bolt distribution. Moreover, UAV aerial photographs the transmission line bolt maps will be affected by lighting, equipment jitter and other factors. This is not conducive to obtaining a finer feature map in the convolution layer. Therefore, this paper adds spatial attention after

the convolutional layer, and by introducing the cross-attention module twice, it helps the detection model to obtain contextual information of the whole image, capture global dependencies, and regions with similar features enhance each other, making the bolts stand out in the full field of view and helping the convolutional layer to obtain a feature map with more obvious features.

In order to help the model obtain global contextual information, this paper introduces the Recurrent Criss-Cross Attention (RCCA) module (Huang et al., 2019) to help feature enhancement of regional features with similar characteristics in a more efficient way. In this paper, we choose the Spatial Attention module, which works as shown in Figure 5 and is composed by two cross-cross attentions, as a simple and efficient way to perform feature enhancement.

Firstly, local features are transmitted through a cross attention module to collect contextual information in both horizontal and vertical directions. Then, by inputting the feature map generated by the first cross attention module into another module, the additional contextual information obtained from the cross path ultimately enables each pixel to capture the full image dependency relationship.

The structure diagram of the cross-attention module is shown in Figure 6. The input feature map is $A_3 \in R^{C \times W \times H}$, which is divided into three branches: Q, K, and V. The feature maps Q and K are obtained through the convolution operation of the 1×1 convolution kernel, where $\{Q, K\} \in R^{C^* \times W \times H}$, then the attention map $A \in R^{(H+W-1) \times W \times H}$ is obtained through Formula 1 and softmax layer.

$$d_{i,u} = Q_u M_{i,u}^T \quad (1)$$

where $d_{i,u} \in D$ represents the weight of the relation between Q_u and $M_{i,u}$, $D \in R^{(H+W-1) \times W \times H}$. $Q_u \in R^{C^*}$ is the value of u position in the spatial dimension of the feature graph Q, $M_u \in R^{(H+W-1) \times C^*}$ is the set of peer or same column elements of u position on K, therefore $M_{i,u} \in R^{C^*}$ is the i th element in M_u .

The other branch V is obtained by 1×1 convolution $V \in R^{C \times W \times H}$, and a new feature map is obtained between V and attention diagram by Formula 2. $V_u \in R^C$ is the value of u position on the spatial dimension of the feature graph V, and $N_u \in R^{(H+W-1) \times C}$ is the set of peer or same column elements of u position on V.

$$N_u^* = \sum_{i=0}^{H+W-1} A_{i,u} N_{i,u} + H_u \quad (2)$$

where N_u^* is the feature vector of position u in $N^* \in R^{C \times W \times H}$, $A_{i,u}$ is the i th value corresponding to the position u in the attention diagram A. Finally, H^* is output in the form of residual error, which enhances the pixel-level expression ability, aggregates the global context information, and improves the performance of bolt and its defect target detection.

3.2 Positional reasoning

Bolts on transmission line coupling plates generally have a fixed position, geometrically in a triangle, and existing target detection models are only for individual targets, with little attention paid to the positional geometric relationships between targets. In this paper, we choose a positional relationship inference module to improve the detection accuracy of the model by using the fixed position information of the bolts on the coupling plates. This is done by using the feature $Q = \{q_i\}$ as an input describing the geometric features of each region to capture the spatial knowledge of the

target (Hu et al., 2018). The position relationship inference module integrates inference from M regions to regions, which is constructed into M region graphs $G(V, E)$ by stacking M MLPs, and uses the edge E to combine the region features with the region spatial location V to learn the position relationship between the regions.

First, the location features q_i of region i and q_j of region j are extracted separately, and the edge values of the two regions are learned using MLP:

$$\hat{e}_{ij} = \text{MLP}(a(q_i, q_j)) \quad (3)$$

where $a()$ indicates the difference between the two regions.

By stacking M MLPs to obtain M location-relative area maps, the edge values of the M location-relative area maps are accumulated and averaged, and they are summed with the unit matrix I to obtain the edge connections:

$$e_{ij} = \frac{1}{M} \sum_{m=1}^M \hat{e}_{ij} + I \quad (4)$$

Afterwards, the location and area features are connected using matrix multiplication to obtain the enhanced features F_s :

$$F_s = \varepsilon F W_s \quad (5)$$

where $\varepsilon \in R^{N \times N}$ is the set of edges of the location-relative region graph, $e_{ij} \in \varepsilon$, F is the input region feature, and W_s is the transformation weight matrix.

4 Experimental design and result analysis

In this paper, the State Grid's "Specification for Image Labeling of Defects in Overhead Transmission Line Equipment (Trial)" and the PASCAL VOC (Everingham et al., 2010) dataset construction method are referred to when constructing the dataset, and the data are annotated in strict accordance with the annotation specification. The dataset is mainly based on a large number of inspection images obtained by UAV inspection with image acquisition equipment, and the inspection image library is filtered and optimized according to manual empirical knowledge, using the joint board target as the main target, providing important data support for the construction of the bolt and its defect detection database.

This paper uses the widely used Precision (P), Recall (R), Intersection over Union (IoU), Average Precision (AP) and mean Average Precision (mAP) in the field of object detection as an evaluation indicator for the accuracy of bolts and their defective targets.

P, R, and IoU are defined in Eqs 6–8 respectively. AP_i^t is the accuracy of the target in category i at an IoU threshold of t , and is defined in Eq. 9 as the value of the area bounded by the Precision-Recall (PR) curve and the coordinate axis. The final mAP is the average of the accuracies of all classes at 10 different IoU thresholds and is used to assess the overall accuracy of the model, which is defined in Eq. 10.

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$IoU = \frac{S_{pre} \cap S_{gt}}{S_{pre} \cup S_{gt}} \quad (8)$$

$$AP_i^t = \int_0^1 P_i(R_i) dR \quad (9)$$

$$mAP = \frac{1}{10} \sum_{t \in \{0.5, 0.55, \dots, 0.95\}} \frac{\sum_{i=1}^C AP_i^t}{C} \quad (10)$$

where TP denotes correct positive samples, FP denotes incorrect positive samples, TN denotes correct negative samples, FN denotes incorrect negative samples, S_{pre} denotes detection results, S_{gt} denotes actual results, i denotes the i th category of bolts or defects, t denotes the threshold of IoU, and C denotes the total number of all categories.

In this paper, transmission line bolts and their defective datasets are selected for experimentation, containing three categories of the normal bolt, pin losing and nut losing, with the specific number of labels for each category shown in Figure 7, with a total of 340 images. It can be seen that the transmission line bolts and their defects dataset constructed in this paper has fewer defective samples and more normal samples, showing a serious long-tail distribution, which is in line with the current general *status quo* of more normal samples and fewer defective samples for transmission line bolt components.

4.1 Comparative experiment on improvement methods

In order to verify the effectiveness of the method in this paper, experiments were carried out using Faster R-CNN as the baseline model and ResNet-101 as the backbone network. The commonly used evaluation metrics in target detection models, mAP, as well as AP, were selected to evaluate the model. The detection results before and after adding the spatial attention module and position relation inference are shown in Table 1. It can be seen that the detection of bolts and their defects by the method in this paper is significantly better than the traditional Faster R-CNN detection model, which does not consider the spatial context information in the detection of bolts and their defects, and only detects the bolts themselves without inference capability. The original Faster R-CNN detection model does not consider the spatial context information in the detection of bolts and their defects. Therefore, this paper adds a spatial attention module and a position relationship inference module to the Faster R-CNN detection model. The improved model improves the mAP by 6.61%, which significantly improves the detection accuracy of the model for transmission line bolts and their defect dataset, which shows the superiority of the proposed improvement.

Table 1 also gives the results of AP values for the baseline model with Faster R-CNN as the detection model and ResNet-101 as the backbone network, the baseline model with spatial attention added, the baseline model with positional relationship inference added, and the method in this paper. It can be seen that adding spatial attention to the baseline model can effectively enhance the label features, especially for the nut losing feature which is difficult to be detected by the baseline model, the feature enhancement effect of spatial attention is obvious, and the AP value of nut losing is improved by 12.83%.

The addition of the location relationship inference module to the baseline model can also effectively improve the detection effect of the model. Especially for the pin losing and nut losing with a small number of labels, the AP values increased by 4.35% and 9.69% respectively. This indicates that the location relationship inference module can effectively exploit the location relationship between bolts. This improves the detection capability of the model, alleviates the long tail effect of the dataset and helps improve the detection accuracy of targets with fewer labels. Finally, the baseline model with Faster R-CNN as the detection model and ResNet-101 as the backbone network is added to the proposed method with the spatial attention module and the position relationship inference module respectively. It can be seen that, on top of the baseline detection model, the AP values of the three types of tags, namely, normal bolt, pin losing and nut losing, increase in this paper method, where the baseline detection The nut losing, which has a lower accuracy, has increased by 13.63% and the mAP has increased by 6.61%.

It can be seen that the location relationship-based transmission line bolt and its defect detection algorithm proposed in this paper are very effective in detecting the state of the bolt. The spatial attention module is used to effectively extract global context information and capture global dependencies, which can effectively help the model extract finer and more comprehensive features. The location relationship inference module is used to add detection box location information to the detection model, which gives the model certain inference capabilities. This method enables the model to use tags with good detection effects and numerous tags, helping to improve detection accuracy for tags with poor detection effects and a small number of tags. Through the above methods, the overall detection ability of the model is improved. In addition, this paper also adds spatial attention and location relationship inference modules to the Faster R-CNN detection model with VGG16 as the backbone network, and the experimental results are shown in Table 2. The results show that the detection effect of the model with VGG16 as the backbone network is significantly worse than that of the model with ResNet-101 as the backbone network. This is because VGG16 performs poorly in feature extraction in the detection of transmission line bolts and their defects, and is unable to extract fine image features for the bolts and their defects. Therefore, the spatial attention module of the method in this paper effectively helps the feature extraction network to perform feature enhancement. When the Faster R-CNN detection model replaces the backbone network, the improvements in this paper can still help the Faster R-CNN detection model to improve its detection capability, especially for the defect samples with few samples and difficult labels to identify, the improvements in this paper can significantly improve its detection accuracy, for example, the AP value of pin losing in this dataset, For example, the AP value of pin losing in this dataset increased by 8.62%, the AP value of nut losing increased by 48.45%, and the total mAP increased by 18.83%.

4.2 Comparative experiment between this method and other methods

In order to further verify the effectiveness of the proposed method for detecting bolts and their defects, experiments were conducted to compare the proposed method with state-of-the-art target detection methods on the same dataset of fixtures, and the methods conducted for comparison included RetinaNet (Lin et al.,

2020), Faster R-CNN, Cascade R-CNN (Cai and Vasconcelos, 2018), and Sparse R-CNN (Sun et al., 2021). As shown in Table 3, the experimental results show that the method in this paper has a higher accuracy for target detection of bolts and their defect datasets.

Table 3 presents several comparative methods, and without considering the computational effort, this method has a significant improvement in accuracy compared to other detection methods, with the mAP of 67.8%. In Table 3, the accuracy of this method is 9.4% higher than that of the single-stage detection method RetinaNet, 6.6% higher than that of the two-stage detection method Faster R-CNN, 2.8% higher than that of the multi-stage detection method Cascade R-CNN, and 0.5% higher than that of Sparse R-CNN, so this method The performance of this paper is better than other detection methods to a certain extent. The experimental data in Table 3 demonstrates that the detection accuracy of this method is generally higher than that of the comparative detection methods and has some practical value.

As shown in Figure 8, several sets of images of the detection results of bolts and their defects on the coupling plate from different shooting angles. Among them, (A), (C), (E), and (G) are the baseline detection results, and (B), (D), (F), and (H) are the detection results of the proposed method. There are four sets of test results in total. In the first set of results, the detection accuracy of both normal bolt and pin losing increased, and the confidence of individual labels increased up to 3.5%. Moreover, the detection frame of each label in this paper fits the label better. In the second set of detection results, the Faster R-CNN detection model failed to detect the difficult samples in the lower left corner, while the proposed method detected the normal bolts in the lower left corner. This is because the spatial attention module added in this article can effectively assist the model in detection, so that smaller targets will not miss detection. In the third set of detection results, both Faster R-CNN and the proposed method detected three labels, but the confidence level of each detection frame of the proposed method is higher, which indicates that the proposed method has a better detection ability. In the fourth set of detection results, the tail of the bolt appears in the Faster R-CNN detection image. While in the actual annotation process, the tail of the bolt is not involved in the annotation. However, Faster R-CNN incorrectly detects the tail of the bolt as a losing pin bolt, and the proposed method successfully avoids such misjudgment.

5 Conclusion

In order to accurately detect transmission line bolts and their defects, this paper uses a joint board to construct a dataset of bolts and their defects, and further conducts the detection of normal bolts, pin losing and nut losing on the fixture. To address these problems of small bolt targets, low image resolution and lack of inference capability of the detection model, a Faster R-CNN detection model based on location relationship inference is used for experimental validation on the self-built dataset. It is demonstrated that adding a spatial attention module after the feature extraction network can effectively help the model enhance the global context information and improve the feature extraction ability of the model; adding a location relationship inference module after the region suggestion can increase the inference ability of the model, help the dataset alleviate the long-

tail effect and improve the AP value of the category with a small number of labels. 6.61%, effectively improving the accuracy of bolt and its defect detection, and laying a good foundation for the task of transmission line bolt and its defect detection.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

Author contributions

ZZ: Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing–review and editing. JX: Conceptualization, Methodology, Validation, Visualization, Writing–original draft. YH: Writing–original draft, Writing–review and editing. SM: Supervision, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research is supported

by the National Natural Science Foundation of China (61871182, U21A20486); Supported by Natural Science Foundation of Hebei Province (F2020502009, F2021502008, F20211502013); Supported by the Fundamental Research Funds for the Central Universities (2023JC006).

Acknowledgments

Heartfelt thanks to everyone who contributed to this paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Cai, Z., and Vasconcelos, N. (June 2018). "Cascade R-CNN: delving into high quality object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (Salt Lake City, UT, USA, 6154–6162. doi:10.48550/arXiv.1712.00726
- Everingham, M., Gool, V., Williams, C., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* 88 (2), 303–338. doi:10.1007/s11263-009-0275-4
- Feng, M., Luo, W., Yu, L., Zhang, P., Hao, X., Fan, Q., et al. (2018). A bolt detection method for pictures captured from an unmanned aerial vehicle in power transmission line inspection. *J. Electr. Power Sci. Technol.* 33 (4), 135–140. doi:10.3969/j.issn.1673-9140.2018.04.019
- Fu, Q., Liu, J., Zhang, X., Zhang, Y., Ou, Y., Jiao, R., et al. (2023). A small-sized defect detection method for Overhead transmission lines based on convolutional neural networks. *IEEE Trans. Instrum. Meas.* 72, 1–12. doi:10.1109/TIM.2023.3298424
- Ge, W., Luo, H., Zhou, G., Fan, B., and Ma, Y. (June 2017). "The application of internet of things technology in power transmission line condition monitoring system," in Proceedings of the International conference on fuzzy systems and data mining (Hualien, Taiwan: IOS Press BV), 485–493. doi:10.3233/978-1-61499-828-0-485
- Gong, Y., Zhou, W., Wang, K., Wang, J., Wang, R., Deng, H., Liu, G., et al. (2023). Defect detection of small cotter pins in electric power transmission system from UAV images using deep learning techniques. *Electr. Eng.* 105, 1251–1266. doi:10.1007/s00202-022-01729-8
- Hu, H., Gu, J., Zhang, Z., Dai, J., and Wei, Y. (June 2018). "Relation networks for object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (Salt Lake City, UT, USA: IEEE), 3588–3597. doi:10.48550/arXiv.1711.11575
- Huang, Z., Wang, X., Wei, Y., Huang, L., Shi, H., Liu, W., et al. (2019). CCnet: criss-cross attention for semantic segmentation. *IEEE/CVF Int. Conf. Comput. Vis. Seoul, South Korea* 45 (6), 6896–6908. doi:10.1109/TPAMI.2020.3007032
- Lin, T., Goyal, P., Girshick, R., He, K., and Dollár, P. (2020). Focal loss for dense object detection. *IEEE Trans. Pattern Analysis Mach. Intell.* 42 (2), 318–327. doi:10.1109/tpami.2018.2858826
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., and Fu, C. (2016). "Ssd: single shot multibox detector," in *European conference on computer vision* (Amsterdam, Netherlands: Springer), 21–37. doi:10.1007/978-3-319-46448-0_2
- Luo, P., Wang, B., Wang, H., Ma, F., Ma, H., and Wang, L. (2023). An ultrasmall bolt defect detection method for transmission line inspection. *IEEE Trans. Instrum. Meas.* 72, 1–12. doi:10.1109/TIM.2023.3241994
- Qi, Y., Wu, X., Zhao, Z., Shi, B., and Nie, L. (2021). Bolt defect detection for aerial transmission lines using Faster R-CNN with an embedded dual attention mechanism. *J. Image Graph.* 26 (11), 2594–2604. doi:10.11834/jig.200793
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *IEEE conference on computer vision and pattern recognition* (Las Vegas, NV, USA), IEEE, 779–788.
- Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Analysis Mach. Intell.* 39 (6), 1137–1149. doi:10.1109/TPAMI.2016.2577031
- Sheng, G., Qian, Y., Luo, L., Song, H., Liu, Y., and Jiang, X. (2021). Key technologies and application prospects for operation and maintenance of power equipment in new type power system. *High. Volt. Eng.* 47 (9), 3072–3084. doi:10.13336/j.1003-6520.hve.20211258
- Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., et al. (June 2021). "Sparse R-CNN: end-to-end object detection with learnable proposals," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (Nashville, TN, USA, 14449–14458. doi:10.1109/CVPR46437.2021.01422
- Zhang, S., Wang, H., Dong, X., Li, Y., Li, Y., and Wang, X. (2021). Bolt detection technology of transmission lines based on deep learning. *Power Syst. Technol.* 45 (7), 2821–2829. doi:10.13335/j.1000-3673.pst.2020.1336
- Zhao, Z., and Ding, J. (2022). Weakly supervised detection method for pin-missing bolt of transmission line based on Improved PCL model. *Sci. Technol. Eng.* 22 (23), 10169–10178. doi:10.3969/j.issn.1671-1815.2022.23.035
- Zhao, Z., Jin, C., Qi, Y., Zhang, K., and Kong, Y. (2021). Image classification of transmission line bolt defects based on dynamic supervision knowledge distillation. *High. Volt. Eng.* 47 (2), 406–414. doi:10.13336/j.1003-6520.hve.20200834
- Zhao, Z., Qi, H., Qi, Y., Zhang, K., and Zhao, W. (2020). Detection method based on automatic visual shape clustering for pin-missing defect in transmission lines. *IEEE Trans. Instrum. Meas.* 69 (9), 6080–6091. doi:10.1109/TIM.2020.2969057
- Zhuo, Z., Zhang, N., Kang, C., Jiang, W., and Wang, Z. (2023). Quantitative attribution analysis method of power system planning scheme for carbon emission peak and carbon neutrality goals. *Automation Electr. Power Syst.* 47 (2), 1–14. doi:10.7500/AEPS20220110006



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Xialei Zhang,
Shanxi University, China
Qian Lu,
Qingdao University, China

*CORRESPONDENCE

Dou An,
✉ douan2017@xjtu.edu.cn

RECEIVED 06 September 2023

ACCEPTED 23 October 2023

PUBLISHED 03 November 2023

CITATION

Cui F, An D and Zhao Y (2023), A two-stage distributed optimization method for home energy management systems via multi-modal data-driven algorithm.
Front. Energy Res. 11:1289641.
doi: 10.3389/fenrg.2023.1289641

COPYRIGHT

© 2023 Cui, An and Zhao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A two-stage distributed optimization method for home energy management systems via multi-modal data-driven algorithm

Feifei Cui, Dou An* and Yingzhuo Zhao

School of Automation Science and Engineering, Faculty of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China

The home energy management system (HEMS), which utilizes multi-modal data from multiple sensors to generate the knowledge about decision making, is essential to the optimization of home energy management efficiency. Load scheduling based on HEMS can improve the utilization efficiency of multi-modal data and derived knowledge, achieve power supply-demand balance, and reduce users' electricity costs. This paper proposes a distributed load optimization scheduling method for the load scheduling problem in HEMS based on multi-modal data-driven algorithm. Additionally, a two-stage data-driven optimization method is proposed, including a first-stage optimization model based on minimizing electricity costs and a second-stage optimization model based on minimizing system load fluctuations. In the first stage, cost self-optimization is performed based on energy storage devices. In the second stage, a load optimization instruction is issued by the control center, and each user optimizes the load fluctuations based on the system load data. Compared to centralized control methods, this approach reduces the computational overhead of the control center. Finally, simulation experiments based on load scheduling in the HEMS are conducted. The results of the first optimization stage show that when the battery capacity integrated into the system increases from 3.68 kWh to 6.68 kWh, user costs can be reduced from 57.572 cents to 42.064 cents. It is not only evident that the proposed method can effectively save users on electricity costs, but the introduction of larger capacity batteries also lowers these costs. The second stage of load fluctuation optimization results show that the proposed method can effectively optimize the usage data of a group of users and decrease the absolute peak-valley difference by 8.8%.

KEYWORDS

multi-modal data, home energy management system, knowledge reasoning, distributed optimization, real-time electricity pricing

1 Introduction

1.1 Background

The next-generation smart grid is a network composed of digital systems and electrical infrastructure, capable of detecting multi-modal data from multiple sensors to monitor the status of energy usage. Additionally, the smart grid enables decision making technologies to reason the knowledge from multi-modal data to implement demand response and load dispatching functions, quickly and adaptively adjusting power generation and transmission (Mahela et al., 2022). The residential load is an essential part of the multi-modal data in smart grids, and as the application of residential load facilities increases, the energy consumption of residential loads continues to grow. Therefore, residential energy management is an urgent and crucial field that can enable end-users to actively participate in reshaping demand patterns.

A home energy management system (HEMS) is the product of the combination of smart grid intelligent communication and data-driven decision-making technology (Lin, 2021). HEMSs can effectively monitor household energy consumption through smart sensors and electronic devices and predict and plan home energy usage, thereby improving electricity utilization efficiency. To guide users to participate in demand response, power companies have introduced real-time pricing strategies in place of traditional fixed prices (Li et al., 2019). Real-time pricing plays an essential role in HEMSs, dividing the day into 24 or more time periods, achieving intelligent meter billing based on real-time prices (Wei et al., 2019). Governments and power grids collect detailed, real-time electricity data from customers via smart meters. This aids in balancing power generation with consumption, thereby stabilizing power system operations and reducing long-term production costs. Power companies set higher prices during peak hours and lower prices during off-peak periods, encouraging users to participate in power system operation management through real-time pricing. Users, aiming to reduce their energy costs, tend to use energy-saving appliances and shift their power usage from peak periods to non-peak periods, thereby improving energy efficiency while reducing electricity costs (Munankarmi et al., 2022). To promote two-way communication, advanced metering infrastructure (AMI) is an essential part of the smart grid's HEMS (Lu et al., 2017), consisting of home area network (HAN), building area network (BAN), neighborhood area network (NAN), and other grid infrastructure (Huang et al., 2021).

Load scheduling based on HEMS can reduce energy consumption, save resources, save electricity costs for consumers and utilities, reduce greenhouse gas emissions, and reduce peak electricity demand. For example, Bejoy et al. (2017) proposed a household appliance scheduling method considering customer preferences and satisfaction to minimize energy consumption without causing inconvenience to users. Pal et al. (2017) used household electric vehicles to manage user load demand and proposed a framework including different appliance energy consumption loads, such as basic load, movable appliances, storage systems, and electric vehicle loads.

Although many achievements have been made in load scheduling based on HEMSs, the main method is to optimize

the system load fluctuation by using the centralized control method, and the modeling of electrical equipment is not practical enough. This paper established the basic equipment and energy storage equipment models, and implemented user load scheduling through two-stage distributed data-driven optimization method. The main contributions are as follows:

- A distributed load scheduling framework for HEMS is proposed, and detailed modeling for various devices is conducted. In the distributed scheme, optimization control is decentralized to individual residential buildings or even to each household user, aiming to reduce the computational and communication overhead of the control center. HEMS is employed to facilitate bidirectional communication and distributed optimization.
- A data-driven two-stage optimization method is introduced. It aims to achieve demand response by optimizing for minimal user costs and minimal load fluctuations, respectively. In the first stage of distributed optimization, users' demand response adjustments can create new peaks in the system load curve. users' demand response adjustments can create new peaks in the system load curve. Users within an area share their optimized consumption from smart meters, and the control center releases system load directives. This prompts a secondary optimization where users exchange load data and iteratively adjust schedulable loads and battery statuses until load fluctuations remain within defined limits.
- Load scheduling simulation experiments were conducted, and the influence of battery parameters was analyzed. Simulation results indicate that, using the proposed method, users can adjust the load based on comfort levels and the urgency of device usage. The obtained scheduling strategy can effectively reduce user costs and decrease load fluctuations.

1.2 Research status

Residential users are a highly important component of the smart grid, accounting for 33% of total electricity consumption. Load scheduling based on home energy management systems (HEMSs) implements demand response from the resident side, and the implementation process faces many challenges, such as privacy leakage, randomness and management complexity of generation and consumption equipment. Therefore, researchers have introduced solutions based on energy storage devices (Seal et al., 2023), distributed energy (Chen and Chang, 2023), flexible loads (Yang et al., 2020), etc. Load scheduling based on HEMS can reduce energy consumption, save resources, save electricity costs for consumers and utilities, reduce greenhouse gas emissions, and reduce peak electricity demand.

Bejoy et al. (2017) proposed a household appliance scheduling method that considers customer preferences and satisfaction to minimize energy consumption without causing inconvenience to users. Pal et al. (2017) used household electric vehicles to manage

user load demand and proposed a framework that includes different appliance energy consumption loads, such as basic load, movable appliances, storage systems, and electric vehicle loads. In the optimization process, each user can arrange their devices based on actual electricity usage. The scheme adds a bias cost to the objective function in the user cost minimization problem to prevent the formation of new peaks during non-peak periods, but using centralized control methods increases the computational burden of the system. Jiang and Wu (2020) proposed a cost-efficient load scheduling method considering user electricity efficiency and satisfaction, balancing user cost and preferences through fractional programming. Kou et al. (2019) introduced a distributed control scheme based on aggregators to achieve residential demand response. In this scheme, the power company provides an incentive price to drive power consumption adjustment based on the aggregated load information exchanged between the utility system and users. Wang et al. (2020) proposed a stochastic optimization method to solve the residential load scheduling problem, establishing residential load models, generation cost prediction models, and stochastic optimal load aggregation models. They introduced a set of uniformly distributed scalars to the load aggregation model to avoid load bounce, and experiments proved that this method effectively reduced the system's load peak-average value. Yang et al. (2018) introduced a privacy-aware scheduling model based on rechargeable batteries, introducing coefficients to enable users to balance privacy protection and cost. The model uses the storage and release of energy by rechargeable batteries to flatten the user's electricity curve and discusses the impact of battery capacity on privacy protection effects. Sangswang and Konghirun (2020) integrated solar energy, energy storage, and V2G. This study provided an optimized control method for electric vehicles and household batteries, enhancing the effectiveness of HEMS. Joo and Choi (2017) proposed a two-stage optimization algorithm for energy consumption scheduling in multiple smart homes under distributed energy. However, this study only considered the interests of consumers and overlooked the quality of the electrical grid. Awad et al. (2015) proposed load scheduling privacy protection methods based on rechargeable batteries and the maximum difference method, using the demand response component to keep the electricity curve constant, and proved that fuzzy processing of smart meter values does not affect user billing. Ming et al. (2016) introduced a user-side load scheduling method that considers user satisfaction, achieving demand response and user cost reduction through two-stage optimization, but did not consider the impact of distributed energy on the smart grid's demand response. The model presented in this paper is a nonlinear programming problem, with variables encompassing both binary and continuous types, and it possesses complex constraints. While some traditional optimization algorithms, such as simulated annealing (Li et al., 2022) and particle swarm optimization (Zhao and Li, 2020), exhibit strong global search capabilities when dealing with nonlinear optimization problems, they encounter challenges when addressing mixed variables, multiple constraints, or high-dimensional problems. Genetic algorithms, on the other hand, can handle both discrete and continuous variables, making them suitable for a wide range of intricate optimization challenges. Therefore, this paper employs the genetic algorithm for model optimization.

2 Home energy management system

2.1 Framework of HEMS

The HEMS (Duman et al., 2021), an essential component of the smart grid, is a microgrid system composed of renewable energy generation equipment, energy storage devices, and various common household appliances. HEMS enables residential end-users to actively participate in reducing peak demand and carrying out demand response. Energy use can be shifted to non-peak periods by scheduling household appliances, reducing excessive energy consumption at certain times. Meanwhile, device scheduling operations must consider customer preferences and satisfaction to achieve the best results in energy scheduling optimization, for example, using air conditioning to maintain the indoor temperature within an appropriate range. To ensure the secure transmission of electricity data and costs between users and the smart grid within the HEMS, advanced metering infrastructure (AMI) is the foundation of HEMS. AMI consists of the home area network (HAN), building area network (BAN), neighborhood area network (NAN), and other grid infrastructure such as smart meters (Huang et al., 2021). The framework of the HEMS is shown in Figure 1.

The framework includes smart meters (SMs), gateways (GWs), control center (CC) of the power company, and users connected to the meters. Smart meters (Fekri et al., 2021) act as a home local area network, installed at the user's end. They are connected to sensor devices in the home and collect user power consumption data through smart appliances, uploading it to the local gateway. Users can monitor and optimize energy control of electrical devices through the home main controller, understand energy usage and related data through smart meters, and choose appropriate electricity usage based on this information to enjoy high quality service. The gateway is a powerful entity used not only for relaying but also for data processing. The control center processes user electricity consumption, encrypted electricity costs, and other data sent from the gateway, and updates real-time prices based on total user electricity consumption, carrying out demand response to keep the load within a certain range in the area, thereby ensuring safe and reliable electricity use. During the transmission process of electricity consumption, for electricity cost-related privacy data, both the gateway and control center will carry out signature authentication to ensure data security and integrity. Users can view their billing accounts through the client and may choose to apply for verification.

2.2 Types of HEMS devices

The devices of HEMSs can be classified as basic devices or energy storage devices.

Basic devices are primarily focused on heating, ventilation, and air conditioning (HVAC), as well as washing machines, refrigerators, rice cookers, etc. Basic devices are divided into schedulable and non-schedulable devices. Non-schedulable devices, such as laptops and desk lamps, must meet users' immediate usage needs, so their operation time cannot be controlled; therefore, non-schedulable devices are not modeled.

Energy storage devices can stitch together intermittent renewable energy and enhance the security and stability of the power

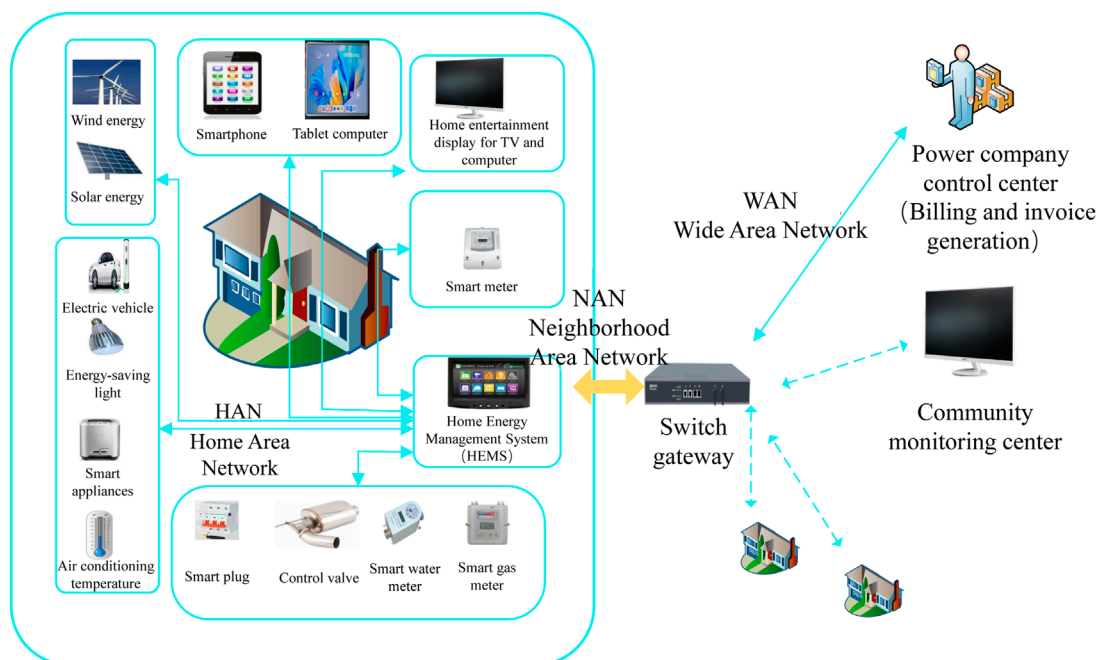


FIGURE 1
Framework of HEMS.

supply system. They can also charge energy storage devices using the grid during off-peak electricity times, store electrical energy through certain mediums, and release the stored energy for power generation during peak electricity periods for household appliance use. This promotes peak shaving and valley filling in the grid, improving the reliability of the user's power supply. Energy storage technologies typically include physical energy storage (flywheel energy storage, pumped hydro storage), chemical energy storage (various types of batteries, renewable fuel cells, supercapacitors), and electrical energy storage (superconducting electromagnetic energy storage). As a load balancing device and backup power source, energy storage systems are also essential equipment for smart grids and distributed energy systems.

3 Distributed system devices model

In the distributed scheme, optimization control is dispersed to individual residential buildings or even individual household users, as shown in Figure 2. Assume that there are W household users in the region, and in a household, in addition to basic electrical devices, energy storage devices are equipped. The home management system is used to provide two-way communication and distributed optimization. In the distributed load scheduling model, the control center is responsible for the publication of real-time electricity prices, system load instructions, and system load data; users can adjust their electricity usage patterns according to different electrical devices in the home, conduct electricity cost self-optimization, and then transmit the smart meter power consumption values to the control center. Furthermore, users transmit system load data to each other to perform system load fluctuation optimization sequentially. Compared to the centralized control scheme, the distributed control

scheme reduces the computational and communication overhead of the control center and provides a scalable architecture.

3.1 Schedulable devices model

Basic devices in the home, such as computers, incandescent lights, televisions, and other appliances that users need to use at any time, are considered non-schedulable devices. Adjusting their usage time would seriously affect users' comfort, so they are not involved in load control. Schedulable devices, such as refrigerators and dishwashers, are referred to as having flexible loads. They can participate in demand response, and their flexibility can alleviate the strain on the power grid during peak electricity usage periods.

In this paper, a day is divided into H intervals, where $h = 1, 2, \dots, H$. The length of each interval is $\Delta h_{\text{step}} = \frac{24 \times 60}{H}$ minutes. $a \in A = \{1, 2, \dots, A\}$ represents the numbers of the electrical devices. Variable $s_a(h)$ represents the working status of device a : when $s_a(h) = 1$, the device is in the working state; when $s_a(h) = 0$, the device is in the off state. $[\alpha_a, \beta_a]$ indicates the permissible working time range for the device, and d_a represents the prescribed working duration of the device. The allowable working time length should be greater than the device's working duration to ensure that the working time of the schedulable device can be rescheduled. Device a should meet the following time constraints:

$$\sum_{h=\alpha_a}^{\beta_a} s_a(h) = d_a \quad (1)$$

$$s_a(h) = 0, \quad h \in H \setminus [\alpha_a, \beta_a] \quad (2)$$

To make the model closer to the actual situation, consideration is given to subdividing the devices.

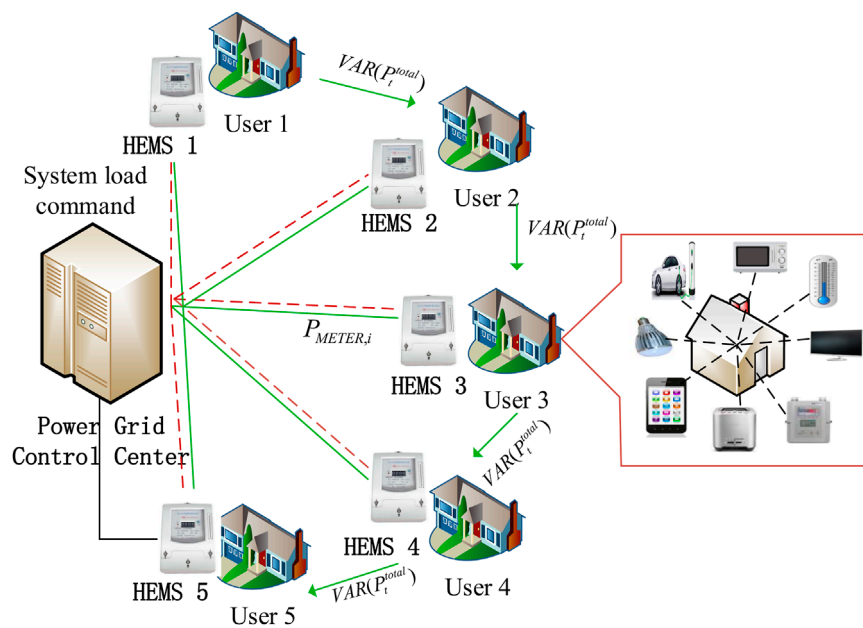


FIGURE 2
Distributed optimization-based load scheduling framework.

3.1.1 Non-interruptible devices

Among the schedulable appliances, devices such as washing machines and rice cookers are considered non-interruptible devices. During operation, they are continuously powered by distributed energy or the grid, and once started, they cannot be stopped, as this would affect the normal functioning of the device. Therefore, once turned on within the schedulable range, they must continue for the specified working duration to complete the corresponding tasks. In addition to satisfying Eq. 1 and Eq. 2, they must meet the following time constraints:

$$\sum_{\tau=h+1}^{h+d_a} s_a(\tau) \geq d_a [s_a(h+1) - s_a(h)] \quad (3)$$

3.1.2 Interruptible devices

Interruptible appliances require intermittent power supply from distributed energy or the grid. Each power supply duration should not be less than the minimum supply time (typically the minimum interval is 30 min or 15 min). With the condition of meeting the minimum interval, these devices can be turned on or off at any time. Examples include microwaves and air conditioners. In addition to satisfying Eq. 1 and Eq. 2, they must meet the following time constraints:

$$P_c(h) = \sum_{a=1}^A s_a(h) \cdot P_a \quad (4)$$

3.1.3 Constant power devices

Due to the significant proportion of HAVC equipment, such as air conditioners, in household electricity consumption, its power varies continuously over time. In contrast, appliances such as refrigerators generally operate at their rated power. Therefore, they are modeled separately. Assume that when a device starts, the power

is P_a , and when idle, the power is 0. The power of a constant power device is given by:

$$P_c(h) = \sum_{a=1}^A s_a(h) \cdot P_a \quad (5)$$

3.1.4 Power-adjustable devices

Power-adjustable devices, such as temperature-controlled appliances like air conditioners, have power needs that vary continuously and are related to the outdoor temperature. The input parameter for the air conditioner is the day-ahead outdoor temperature, and its mathematical model is represented as:

$$T_{in}(h+1) = \varepsilon T_{in}(h) + (1-\varepsilon) \cdot \left(T_{out}(h) \pm A \frac{P_{NC}(h)}{\eta} \right) \quad (6)$$

where $T_{in}(h)$ is the indoor temperature of time slot h ; ε is the inertia coefficient of the indoor temperature change; $T_{out}(h)$ is the outdoor temperature of time slot h ; A is the thermal capacity of the room; $P_{NC}(h)$ is the rated power of the air-conditioning appliance in time slot h ; η is the thermal conductivity efficiency of the room.

When the air conditioner operates in cooling mode, the value of \pm in the formula is set to $-$; when the air conditioner operates in heating mode, the value of \pm in the formula is set to $+$. Considering user comfort, the indoor temperature should be maintained within the range of user demand, and air-conditioning appliances must satisfy the following constraints:

$$T_{in}^{\min} \leq T_{in}(h) \leq T_{in}^{\max} \quad (7)$$

where T_{in}^{\min} is the minimum indoor temperature set by the user; T_{in}^{\max} is the maximum indoor temperature set by the user. In addition, the power consumption during operation of air-conditioning appliances should satisfy the following constraints:

$$0 \leq P_{NC}(h) \leq P_{NC}^{\max} \quad (8)$$

where P_{NC}^{\max} is the maximum power consumption when the appliance is operating, determined by the nameplate value.

3.2 Energy storage devices model

Energy storage devices can store electrical energy through a certain medium, acting as a buffer between power generation and consumption. This enables users to charge and store energy during off-peak periods and utilize battery-released energy during peak periods, enhancing electricity safety and stability. The primary energy storage device used in homes is the lithium battery.

The main parameters affecting battery operation include: Capacity, State of Charge (SOC) and Charging or Discharging Power of the Battery.

3.2.1 Capacity

Capacity refers to the quantity of electrical charge a battery can store. It is denoted by E_{batt} and is measured in ampere-hours, abbreviated as Ah. Generally, the larger the battery volume, the higher its capacity.

The rated capacity refers to the minimum amount of electrical energy released by the battery at 25°C when discharged at a 10-h rate.

The actual capacity represents the energy a battery can output under certain conditions, equivalent to the product of the current and time.

3.2.2 SOC

The state of charge reflects the ratio of the remaining battery charge to the battery's capacity. To extend a battery's lifespan, its state of charge must be considered during operation to ensure it remains within a certain range. When SOC = 0, the battery is depleted, and when SOC = 1, the battery is fully charged.

$$SOC(h) = \frac{E_B(h)}{E_{batt}} \quad (9)$$

where $SOC(h)$ represents the state of charge of the battery in time slot h ; $E_B(h)$ is the remaining charge of the battery in time slot h ; E_{batt} is the battery capacity.

The charge of the battery in time slot h is calculated according to Eq. (10):

$$E_B(h) = E_{B0} + \sum_{\tau=1}^h P_B(\tau) \quad (10)$$

where E_{B0} is the initial charge of the battery; $P_B(h)$ is the charge/discharge power of the battery in time slot h .

The state of charge of the battery is influenced by its charging and discharging, and the dynamic change process is described in Eq. (11):

$$SOC(h+1) = SOC(h) + \frac{(P_B^{ch}(h) - P_B^{dch}(h)) \cdot \Delta h_{step}}{E_{batt}} \quad (11)$$

where $P_B^{ch}(h)$ is the charging power of the battery in time slot h ; $P_B^{dch}(h)$ is the discharging power of the battery in time slot h .

An excessively high or low SOC is detrimental to the battery's lifespan. Therefore, constraints on the SOC range are shown in Eq. (12):

$$SOC^{\min} \leq SOC(h) \leq SOC^{\max} \quad (12)$$

where SOC^{\min} is the minimum allowable state of charge for the battery, SOC^{\max} is the maximum allowable state of charge for the battery. When the battery's state of charge falls below SOC^{\min} , the battery will no longer discharge; when the state of charge exceeds SOC^{\max} , the battery will no longer charge.

3.2.3 Charging or discharging power of the battery

When the battery is in operation, it is either in a charging state or in a discharging state. A 0–1 variable a is introduced to represent the state of the battery. b indicates the battery is in a charging state during time slot h , while c indicates the battery is in a discharging state during time slot h . To extend the battery's lifespan, one cannot arbitrarily switch between charging and discharging states. Therefore, this paper maintains that a state switch can occur only after controlling the charging or discharging state for more than 30 min.

The constraints for the charging and discharging power of the battery in each time slot are shown in Eq. 13 and Eq. 14:

$$0 \leq \frac{P_B^{ch}(h)}{\eta_{ch}} \leq S_B(h) \cdot P_{ch}^{\max} \quad (13)$$

$$0 \leq P_B^{dch}(h) \cdot \eta_{dch} \leq (1 - S_B(h)) \cdot P_{dch}^{\max} \quad (14)$$

where η_{ch} is the charging efficiency of the battery; P_{ch}^{\max} is the maximum amount of electricity allowed to be charged in one time slot; η_{dch} is the discharging efficiency of the battery; P_{dch}^{\max} is the maximum amount of electricity allowed to be discharged in one time slot.

The charging and discharging power of the battery is:

$$P_B(h) = \frac{P_B^{ch}(h)}{\eta_{ch}} - P_B^{dch} \cdot \eta_{dch} \quad (15)$$

In the equation, when the battery is in a charging state, $P_B^{dch}(h)$ is 0, and at this time, the battery's charging and discharging power is the charging power $\frac{P_B^{ch}(h)}{\eta_{ch}}$. When the battery is in a discharging state, $P_B^{ch}(h)$ is 0, and at this time, the battery's charging and discharging power is the discharging power $P_B^{dch} \cdot \eta_{dch}$.

4 Two-stage distributed optimization model

4.1 First-stage optimization model

Each user household engages in flexible load scheduling, autonomously choosing their electricity consumption time. They opt to use electricity during low tariff periods, ensuring their electricity needs are met and thereby reducing household electricity costs. In the model of this paper, the smart grid can exchange electricity bidirectionally with users. That is, users can sell their excess energy

back to the main grid. The optimization objective for minimizing electricity costs is expressed as:

$$\min \text{Cost}_{elec} = \min \left\{ \sum_{h=1}^H RTP(h) \cdot P_{METER}(h) \right\} \quad (16)$$

where $RTP(h)$ is the real-time electricity price published by the power grid company; $P_{METER}(h)$ is the power consumption value recorded by the smart meter during time slot h .

In a Grid-Feeding HEMS, $P_{METER}(h)$ can take both positive and negative values. When $P_{METER}(h)$ is positive, the household is purchasing electricity from the grid. Conversely, when $P_{METER}(h)$ is negative, the household is feeding electricity back to the grid. Users can obtain real-time electricity prices $RTP(h)$ in advance from the power grid company.

The calculation method of $P_{METER}(h)$ is shown in Eq. (17):

$$\begin{aligned} P_{METER}(h) &= P_{LOAD}(h) + P_B(h) \\ &= P_M(h) + P_C(h) + P_{NC}(h) + P_B(h) \end{aligned} \quad (17)$$

where $P_{LOAD}(h)$ represents the total load of basic household electrical appliances; $P_B(h)$ represents the battery's charging and discharging quantity; $P_M(h)$ is the power consumption of non-dispatchable loads; $P_C(h)$ is the power consumption of power-stable devices among dispatchable loads; and $P_{NC}(h)$ is the power consumption of power-adjustable devices among dispatchable loads.

4.2 Second-stage optimization model

Each household user, in order to save costs, participates in demand response and adjusts their own electricity consumption behavior, which might introduce new peak demand for the system. To ensure the safe and stable operation of the grid and prevent this situation, the model takes into account the collective residential load in a specific region and incorporates system deviation costs into the objective function to minimize. This approach reduces the peak-to-average-power ratio (PAR) and prevents the system from encountering new peaks during non-peak periods.

The optimization objective for minimizing load fluctuation is expressed as:

$$\min \left\{ \gamma \cdot \text{Cost}_{elec} + (1 - \gamma) \cdot \sum_{h=1}^H \left[\sum_{w \in W} P_{METER,w}(h) - \bar{P}_{TOTAL} \right]^2 \right\} \quad (18)$$

where Cost_{elec} represents the user's self-optimized electricity cost; γ is the weight factor of the deviation cost; w refers to a user in household w . The second term of the function, denoted as $\text{VAR}(P_h^{TOTAL})$, is used to evaluate the load fluctuation of the user group.

\bar{P}_{TOTAL} is calculated according to Eq. (19):

$$P_h^{TOTAL} = \sum_{w \in W} P_{METER,w}(h) \quad (19)$$

$$\begin{aligned} \bar{P}_{TOTAL} &= \frac{1}{H} \sum_{h=1}^H (P_h^{TOTAL}) \\ &= \frac{1}{H} \sum_{h=1}^H \left(\sum_{w \in W} P_{METER,w}(h) \right) \end{aligned} \quad (20)$$

P_h^{TOTAL} represents the power consumption of the system in time slot h , which is the total power consumption of W household users. The physical significance of \bar{P}_{TOTAL} is the average power consumption of W household users in the region over H time slots. By controlling the sum of power consumption in each time slot for each household user to be close to \bar{P}_{TOTAL} , the system load fluctuation can be reduced. Different values of γ can be chosen to strike a balance between cost and system load fluctuation.

Final model output:

- 1) The $(A + 3) \times H$ -dimensional flexible load state matrix X_{Chrom} represents the working status of all flexible loads over a 24-h day after participating in load scheduling, as shown in Eq. (21):

$$X_{Chrom} = \begin{bmatrix} X_{S1} \\ X_{S2} \\ \vdots \\ X_{Sa} \\ \vdots \\ X_{SA} \\ X_{SB} \\ X_{PT} \\ X_{PB} \end{bmatrix} = \begin{bmatrix} x_{s1}^1 & x_{s1}^2 & \cdots & x_{s1}^H \\ x_{s2}^1 & x_{s2}^2 & \cdots & x_{s2}^H \\ \vdots & \vdots & \vdots & \vdots \\ x_{sa}^1 & x_{sa}^2 & \cdots & x_{sa}^H \\ \vdots & \vdots & \vdots & \vdots \\ x_{SA}^1 & x_{SA}^2 & \cdots & x_{SA}^H \\ x_{SB}^1 & x_{SB}^2 & \cdots & x_{SB}^H \\ x_{PT}^1 & x_{PT}^2 & \cdots & x_{PT}^H \\ x_{PB}^1 & x_{PB}^2 & \cdots & x_{PB}^H \end{bmatrix} \quad (21)$$

where X_{Chrom} is a matrix composed of X_{Sa} , X_{SB} , X_{PT} and X_{PB} . Matrix X_{Sa} ($a = 1, 2, \dots, A$) represents the working status of device a , with 0 indicating working and 1 indicating idle. Vector X_{SB} represents the working status of the battery. Vector X_{PT} represents the power of adjustable power devices. Vector X_{PB} represents the charge and discharge power values of the battery.

- 2) All flexible load working states multiplied by the rated power of the corresponding time slot result in the power consumption of the adjustable device for each time slot in a day. This is referred to as the power consumption matrix P_{Chrom} , as shown in Eq. (22):

$$\begin{aligned} P_{Chrom} &= X_{Chrom}^a \times P_a \\ &= \begin{bmatrix} x_{s1}^1 \cdot P_1 & x_{s1}^2 \cdot P_1 & \cdots & x_{s1}^H \cdot P_1 \\ x_{s2}^1 \cdot P_2 & x_{s2}^2 \cdot P_2 & \cdots & x_{s2}^H \cdot P_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{sa}^1 \cdot P_a & x_{sa}^2 \cdot P_a & \cdots & x_{sa}^H \cdot P_a \\ \vdots & \vdots & \vdots & \vdots \\ x_{SA}^1 \cdot P_A & x_{SA}^2 \cdot P_A & \cdots & x_{SA}^H \cdot P_A \\ x_{SB}^1 \cdot x_{PB}^1 & x_{SB}^2 \cdot x_{PB}^2 & \cdots & x_{SB}^H \cdot x_{PB}^H \\ x_{PT}^1 & x_{PT}^2 & \cdots & x_{PT}^H \end{bmatrix} \end{aligned} \quad (22)$$

After HEMS load scheduling, the obtained adjustable device state matrix X_{Chrom} and power consumption matrix P_{Chrom} represent the optimal working state collection that satisfies both the device's inherent constraints and user comfort.

5 Load scheduling based on two-stage distributed optimization

5.1 Scheduling process based on two-stage distributed optimization

One commonly used method to reduce user costs and mitigate system load fluctuations through load scheduling is centralized control. Power company control centers process the power consumption values collected from smart meters in a given region, thereby decreasing system peak averages and smoothing the system load curve. However, centralized control methods present certain challenges. The computational burden at the control center, coupled with the communication overhead from each user transmitting to the center, is considerable. This is primarily because the load state matrix X_{Chrom} of $(A + 3) \times H$ dimension for each user in the region forms a three-dimensional matrix $(A + 3) \times H \times W$, where W is the number of users. All require optimization computations through the control center. As the number of users grows, the computational scale of the aforementioned model substantially increases. Employing distributed optimization control methods can circumvent the curse of dimensionality. Additionally, distributed optimization methods offer robust user privacy protection. Users only need to upload the post-optimization smart meter values, with each household independently optimizing load fluctuations. The data transmitted between users pertains to system load, negating the need for individual household power data. In contrast, centralized control mandates not just the uploading of smart meter consumption data, but also the power and status of each appliance in a household. By readjusting the power and status of appliances, the control center minimizes system load fluctuations. In doing so, it gains access to granular user consumption data, which inevitably breaches user privacy.

Distributed optimization facilitates a layered, phased approach to the optimization process. In this paper's distributed optimization load scheduling model, the first phase encompasses users self-optimizing for cost. Under the premise of ensuring user comfort, the electricity usage time of flexible loads is adjusted to minimize each household's electricity cost. Subsequently, the optimized smart meter consumption values are uploaded. Some of the literature has explored the gradual processing of smart meter consumption values to better safeguard user privacy. The second phase focuses on optimizing system load fluctuations. The control center processes the collected regional smart meter consumption values to obtain aggregate area electricity consumption and system load fluctuation data. The power company's control center then releases system load optimization command $\min VAR(P_t^{TOTAL})$ and dispatches system load fluctuation data \bar{P}_{TOTAL} , initiating the HEMS optimization process. During the regulatory process, load data are transferred among users. Initially, User 1 undergoes electricity optimization through Eq. (18), altering the operational status of interruptible appliances and overall electricity consumption behavior, before relaying the post-optimization system load data to User 2. This sequential process continues for W users, ceasing optimization once the results align with predetermined criteria. The data shared among users are system load data, offering a degree of user privacy protection. The distributed control procedure is depicted in Figure 3.

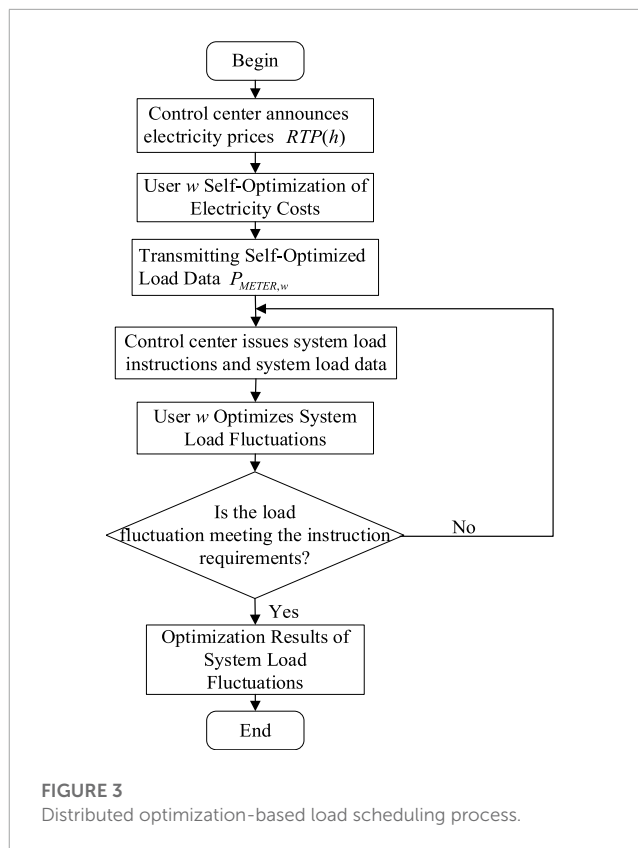


FIGURE 3

Distributed optimization-based load scheduling process.

5.2 Model resolution based on hybrid coding genetic algorithm

Eq. (18) is a nonlinear programming problem. Typical optimization model solutions can use methods such as simulated annealing or particle swarm optimization. However, due to the uniqueness of variables in this paper's model, as indicated by Eq. 21 and Eq. 22, the variables to be resolved include both binary 0–1 variables and continuous variables. Moreover, it possesses stringent constraint conditions. When using a genetic algorithm, the total number of 1's in a chromosome represents the equipment's operating duration. By controlling the positions of 1's in the chromosome, we can set the equipment start and stop times, thus determining its scheduling range, which is in line with the schedulable model. The genetic algorithm demonstrates superior performance in handling this paper's model, hence its selection.

The use of genetic algorithms to solve optimization problems comprises four main steps: potential solution encoding, initial population gene initialization, fitness function computation, and genetic operations. These operations include selection, replication, crossover, and mutation. In this paper's model, apart from the conventional upper and lower limit constraints, there are some unconventional constraints. All electric devices must adhere to the constraints of Eq. 1 and Eq. 2. Non-interruptible devices must comply with the constraints of Eq. (4), which restricts the number and position of occurrences of 0 and 1 values in genes. Traditional genetic algorithms cannot resolve this problem. Modifications are required for potential solution encoding, population generation, and

crossover mutation, necessitating the use of a hybrid coding genetic algorithm.

5.2.1 Initial population

5.2.1.1 Hybrid encoding

In the HEMS model of this paper, the operating status of schedulable devices is a discrete variable, represented by 1 when in operation and 0 when idle. The battery's working state is also a discrete variable, represented by 1 during charging and 0 during discharging. However, the power of adjustable power devices such as air conditioners and the charging and discharging power of batteries are continuous variables, denoted as X_{PT} and X_{PB} , respectively. Therefore, the chromosome composition of an individual is shown in Eq. (23):

$$X_{Chrom} = \begin{bmatrix} X_{S1} \\ X_{S2} \\ \vdots \\ X_{Sa} \\ \vdots \\ X_{SA} \\ X_{SB} \\ X_{PT} \\ X_{PB} \end{bmatrix} = \begin{bmatrix} x_{s1}^1 & x_{s1}^2 & \cdots & x_{s1}^H \\ x_{s2}^1 & x_{s2}^2 & \cdots & x_{s2}^H \\ \vdots & \vdots & \vdots & \vdots \\ x_{sa}^1 & x_{sa}^2 & \cdots & x_{sa}^H \\ \vdots & \vdots & \vdots & \vdots \\ x_{SA}^1 & x_{SA}^2 & \cdots & x_{SA}^H \\ x_{SB}^1 & x_{SB}^2 & \cdots & x_{SB}^H \\ x_{PT}^1 & x_{PT}^2 & \cdots & x_{PT}^H \\ x_{PB}^1 & x_{PB}^2 & \cdots & x_{PB}^H \end{bmatrix} \quad (23)$$

(A+3)×H

where X_{Chrom} is a chromosome group composed of X_{Sa} , X_{SB} , X_{PT} and X_{PB} . H represents the day divided into H time slots. The binary encoded chromosome X_{Sa} ($a = 1, 2, \dots, A$) denotes the working status of device a . The binary encoded chromosome X_{SB} indicates the working status of the battery. The real-number encoded chromosome X_{PT} is the power of adjustable power devices. The real-number encoded chromosome X_{PB} stands for the charging and discharging power of the battery.

Suppose the initial population size is K and that the length of each chromosome is H . The initial population can then be depicted using a three-dimensional matrix X with a size of $(A + 3) \times H \times K$. $X(:, :, k) = X_{Chrom}^k$, $k = 1, 2, \dots, K$. X_{Chrom} represents the chromosome set of an individual, comprising X_{Sa} , X_{SB} , X_{PT} and X_{PB} . The initial population is typically generated randomly, but it must adhere to the relevant constraints.

X_{Sa} must satisfy the constraints of Eq. 1 and Eq. 2, where the number 1 can only appear between the α_a -th and β_a -th positions in X_{Sa} , and the total count of 1s is equal to d_a . For non-interruptible appliances, Eq. (3) must be satisfied, where gene 1 can only appear continuously between the α_a -th and β_a -th positions. The interruptible appliances satisfy Eq. (4), with gene 1 appearing randomly between the α_a -th and β_a -th positions. The operating power of the air conditioner X_{PT} and the charge-discharge power of the battery X_{PB} must also adhere to their respective upper and lower power limits.

5.2.1.2 Fitness function

The fitness function is used to evaluate an individual's adaptability to its environment, determining the probability of its genes being passed on. This directly affects whether the optimal solution can be found and the convergence speed of the

algorithm. The design should be as simple as possible to minimize computational complexity. To apply the genetic algorithm for solution finding, the problem of maximizing the objective function should be transformed into a minimization problem. In the model presented in this paper, the objective functions F for the two phases of distributed optimization take non-negative values. The fitness function is chosen as the reciprocal of the objective function. In the first phase, where users optimize themselves, the fitness function is taken as the reciprocal of the cost function. In the second phase of load fluctuation optimization, the fitness function is the reciprocal of the variance of system load data. Therefore, the fitness function f can be represented as:

$$f = \frac{1}{F} \quad (24)$$

5.2.2 Genetic operations

5.2.2.1 Selection

After calculating the fitness of all individuals, the selection process determines which individuals will participate in reproduction and pass their genes on to the next-generation. Individuals with a high fitness value have a greater chance of being selected, while those with a low fitness value have a lesser chance. Roulette wheel selection is commonly used for this purpose. The probability P_{xi} of individual xi being selected is calculated according to Eq. (25):

$$P_{xi} = \frac{f_i}{\sum_{j=1}^K f_j}, \quad i = 1, 2, \dots, K \quad (25)$$

where f_i is the fitness of the first individual; K stands for the total number of individuals in the population.

5.2.2.2 Crossover

Crossover, or genetic recombination, involves taking two parent individuals and swapping portions of their chromosomes to produce two new chromosomes, thereby creating new offspring individuals. The crossover probability P_c is typically a random number between 0 and 1. Consequently, the probability of the parent chromosomes being directly copied to the next-generation is $1 - P_c$.

In the model presented in this paper, there are both binary-encoded chromosomes and real-number encoded chromosomes. Accordingly, the crossover method should be chosen to match the respective encoding types.

For binary-encoded chromosomes of basic devices, one or multiple crossover points are selected on the parent chromosomes, followed by a swapping operation. It is crucial to ensure that after crossover, the constraints for non-interruptible devices given by Eq. 3 and for interruptible devices given by Eq. (4) are still satisfied.

For chromosomes corresponding to power-adjustable devices and rechargeable battery power encoded in real numbers, calculations are performed with a random number between 0 and 1 and the parental chromosome. If the parental chromosome is

represented by $\begin{cases} X_1 = (x_1^{(1)}, x_2^{(1)}, \dots, x_m^{(1)}) \\ X_2 = (x_1^{(2)}, x_2^{(2)}, \dots, x_m^{(2)}) \end{cases}$, the gene value of the offspring chromosome $\begin{cases} Y_1 = (y_1^{(1)}, y_2^{(1)}, \dots, y_m^{(1)}) \\ Y_2 = (y_1^{(2)}, y_2^{(2)}, \dots, y_m^{(2)}) \end{cases}$ obtained from

crossover is calculated according to Eq. (26):

$$\begin{cases} y_i^{(1)} = \alpha_i x_i^{(1)} + (1 - \alpha_i) x_i^{(2)} \\ y_i^{(2)} = (1 - \alpha_i) x_i^{(1)} + \alpha_i x_i^{(2)} \end{cases}, \quad i = 1, 2, \dots, m \quad (26)$$

5.2.2.3 Mutation

Mutation refers to the periodic random updating of a gene on a chromosome to refresh the population, exploring unknown areas in the solution space.

In the model of this paper, for binary encoded chromosomes, when a random number is less than the mutation probability P_m , the corresponding chromosome's binary string is flipped. When the original gene value at the mutation point is 0, it is flipped to 1; when the original gene value at the mutation point is 1, it is flipped to 0. For real-number encoded chromosomes, a uniformly distributed random number within the value range replaces it, namely, the uniform distribution method. The calculation method for the mutated gene is based on Eq. (27):

$$x'_k = L_{\min}^k + \beta (L_{\max}^k - L_{\min}^k) \quad (27)$$

where x'_k represents the gene value after mutation; L_{\min}^k is the lower limit of the corresponding variable (power of the adjustable power device) of the chromosome; β is a random value from a 0–1 uniform distribution; and L_{\max}^k is the upper limit of the corresponding variable (power of the adjustable power device) of the chromosome.

The algorithm flow of the genetic algorithm is shown in Figure 4.

6 Simulation verification

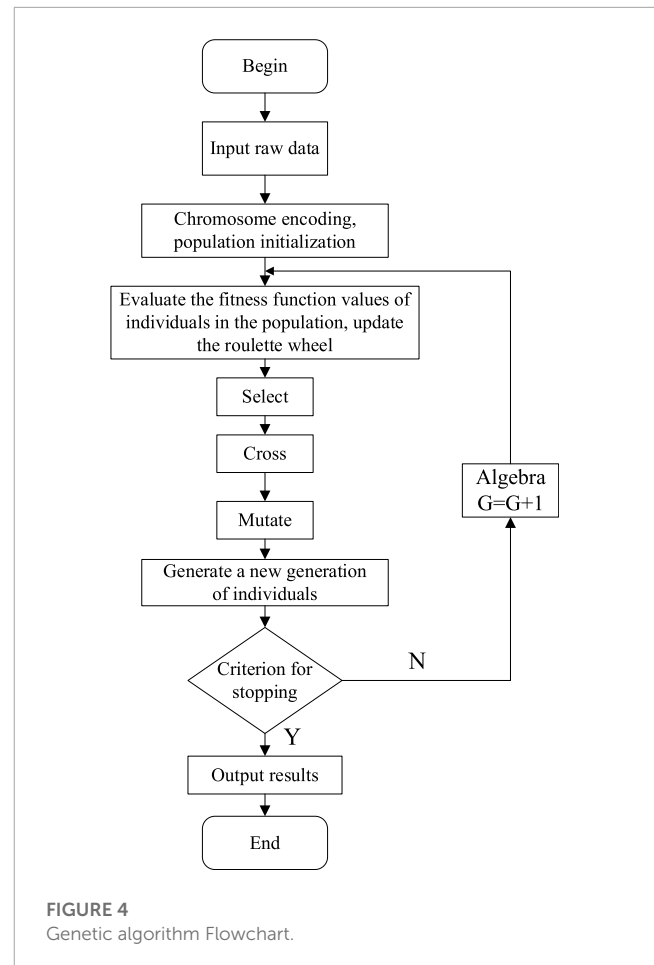
To conduct research on load optimization scheduling for the HEMS, this study designed a distributed optimization load scheduling simulation experiment. In the first phase, users optimize costs for themselves, while in the second phase, the optimization focuses on reducing the system's load fluctuations. The study also investigates the impact of energy storage devices on load scheduling.

6.1 Parameter settings

Dividing the 24-h day into 48 time intervals results in $\Delta h_{step} = 30$ min. Python was used for modeling and solving. The simulation platform was equipped with an Intel(R) Core(TM) i5-10400 CPU at 2.90 GHz, 16 GB of RAM, and ran Windows 10 Home edition as its operating system. The output variables of the experiment are the flexible load status matrix, from which the power consumption matrix of various electrical devices, the consumption values of the smart meter, and the daily electricity costs can be deduced. The data sources and settings are described as follows.

6.1.1 Electricity prices and outdoor temperature data

The outdoor temperature data are taken from the temperature readings of a particular summer day in Xi'an. The electricity prices and outdoor temperature data are shown in Figure 5A, B, respectively.



6.1.2 Parameter settings for dispatchable devices

The primary device chosen for power-adjustable research is the air conditioner. The temperature parameters ε , A and η for the air conditioner are set to 0.93, 2.5, and 0.45, respectively. In the summer, the air conditioner operates in cooling mode, and its maximum allowed power output per time slot is 3.6 kWh. The indoor temperature set by the user must be maintained between 24°C and 26°C.

Power-fixed devices include 20 basic devices, of which device number 2, the washing machine, and device number 16, the rice cooker, are non-interruptible devices. These devices must adhere to the respective time constraints of non-interruptible devices; once activated, they must complete their respective tasks before they can stop. The remaining 18 devices are interruptible. All device tasks are numbered, and the dispatch time range, working duration, and power of dispatchable devices are presented, as shown in Table 1:

6.1.3 Energy storage device parameter settings

The energy storage device selected is a household lithium battery. There are two types of batteries: Battery A has a capacity of 3.68 kWh and a maximum charging/discharging power of 2.5 kW; Battery B has a capacity of 6.68 kWh and a maximum charging/discharging power of 5 kW. The SOC of the battery must be maintained between 0.3 and 0.9.

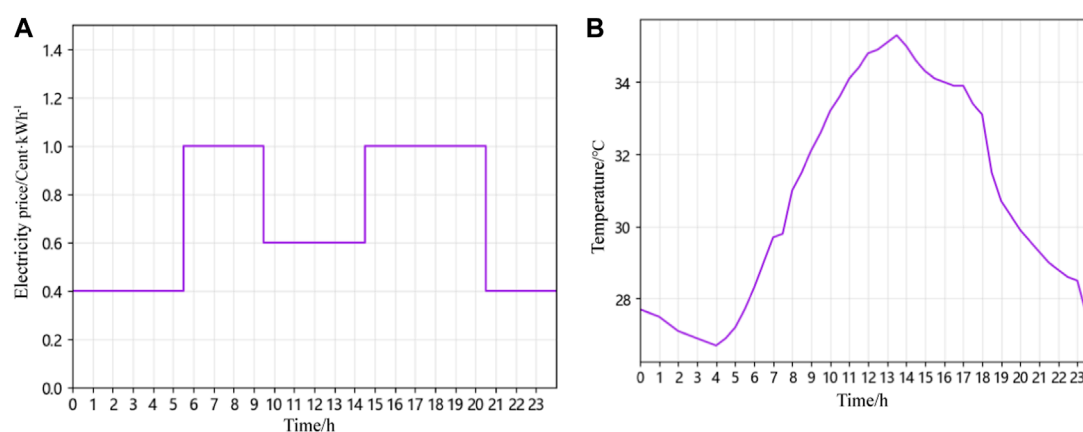


FIGURE 5

Electricity prices and outdoor temperature data: (A) real-time electricity price; (B) outdoor temperature variation curve.

TABLE 1 Parameter settings for dispatchable devices.

Appliance ID	Appliance type	Schedulable area/ $[\alpha_a, \beta_a]$	Working duration/ d_a	Power/kW
1	Range hood or Exhaust hood	10:30:00 a.m.-2:00:00 p.m.	2	0.13
2	Washing machine	12:00:00 a.m.-11:30:00 p.m.	10	1.5
3	Vacuum cleaner	12:00:00 a.m.-11:30:00p.m.	6	0.3
4	Iodine tungsten lamp	8:00:00 p.m.-11:30:00 p.m.	1	0.5
5	Oven	7:00:00 a.m.-11:00:00 a.m.	2	3
6	Water pump	8:00:00 a.m.-11:30:00 p.m.	6	2
7	Microwave oven	7:00:00 a.m.-11:00:00 a.m.	1	1
8	Oil extractor	6:00:00 p.m.-11:30:00 p.m.	5	0.5
9	Air humidifier	5:00:00 a.m.-9:00:00 a.m.	3	0.5
10	Swimming pool pump	11:00:00 a.m.-3:00:00 p.m.	1	2
11	Refrigerator	12:00:00 a.m.-11:30:00p.m.	23	0.5
12	Disinfection cabinet	8:00:00 a.m.-11:30:00 p.m.	4	0.5
13	Printer	8:00:00 a.m.-11:30:00 p.m.	4	0.3
14	Dryer	12:00:00 p.m.-4:00:00 p.m.	4	1.5
15	Electric kettle	7:00:00 p.m.-11:30:00 p.m.	4	1.5
16	Rice cooker	6:00:00 a.m.-6:00:00 p.m.	12	0.5
17	Mixer 1	6:00:00 a.m.-12:00:00 p.m.	2	0.3
18	Mixer 2	6:00:00 a.m.-12:00:00 p.m.	2	0.3
19	Water heater	5:00:00 a.m.-8:00:00 p.m.	12	1.5
20	Hairdryer	5:00:00 a.m.-10:00:00 a.m.	4	0.5

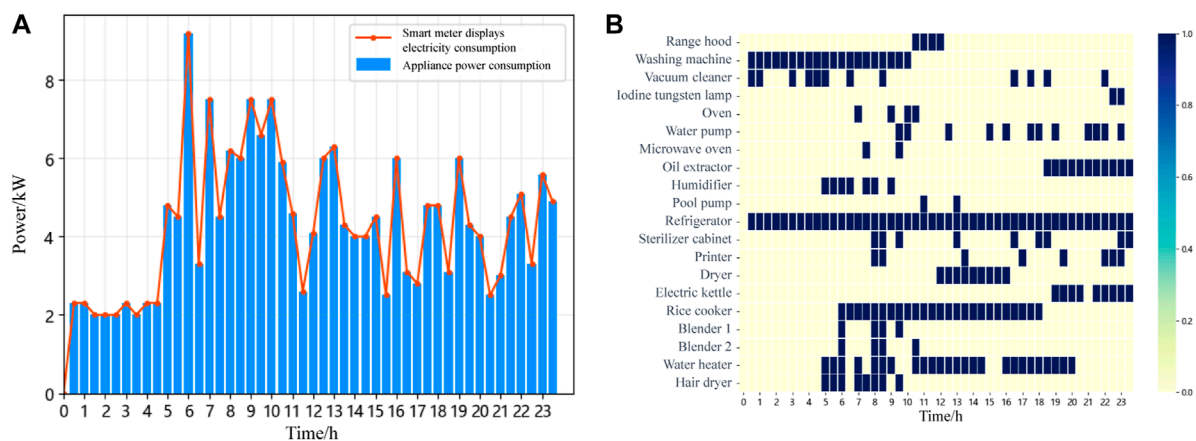


FIGURE 6
Electricity prices and outdoor temperature data: (A) initial electricity load curve; (B) operational status diagram of electrical devices.

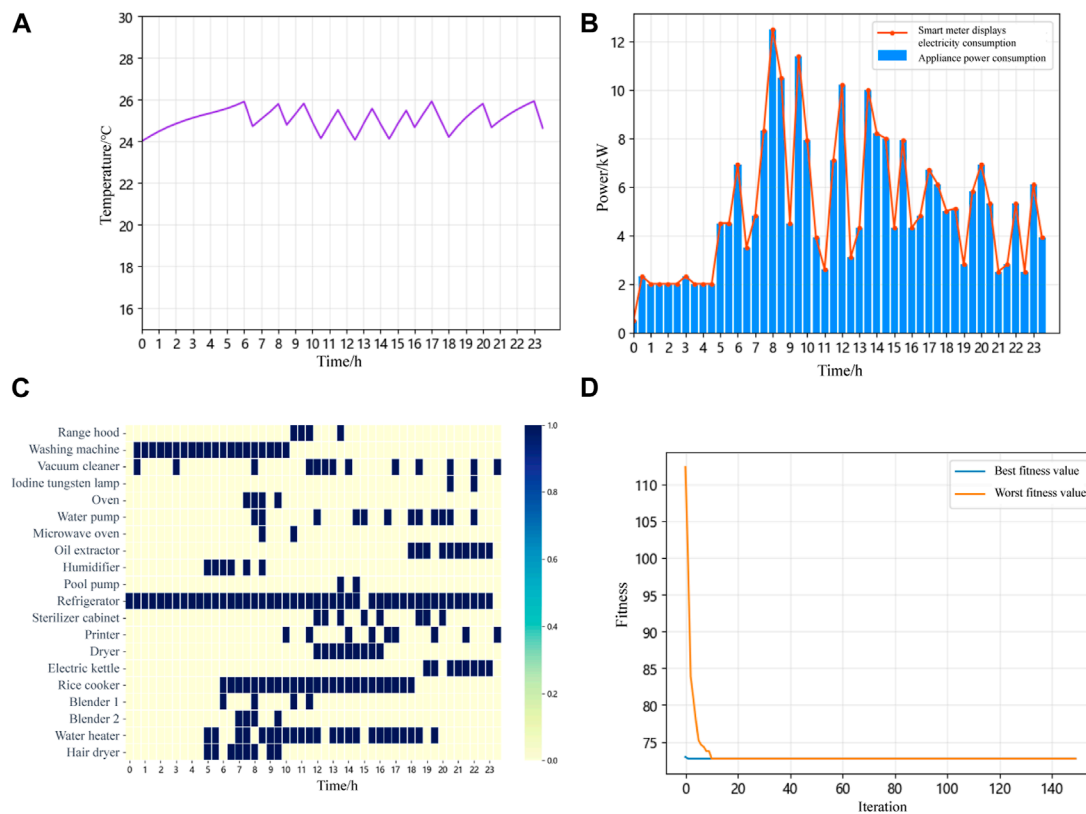


FIGURE 7
Simulation results without energy storage battery integration: (A) indoor temperature variation curve; (B) electricity task scheduling results; (C) operational status diagram of electrical devices; (D) convergence curve of genetic algorithm.

6.1.4 Genetic algorithm parameter settings

In the initialization parameters of the genetic algorithm, the population size $N = 50$. Thus, the initial population can be represented by a three-dimensional matrix X (size: $23 \times 48 \times 50$). The crossover probability $P_c = 0.8$, mutation probability $P_m = 0.1$, and the maximum number of iterations is 150 times.

6.2 First-stage simulation results

The original electricity load curve of a household user within 48 time slots in a day is shown in Figure 6A. At this time, the electricity cost is 115.566 cents. As can be seen from the figure, the user's electricity consumption is concentrated in 6:00–10:30. A significant

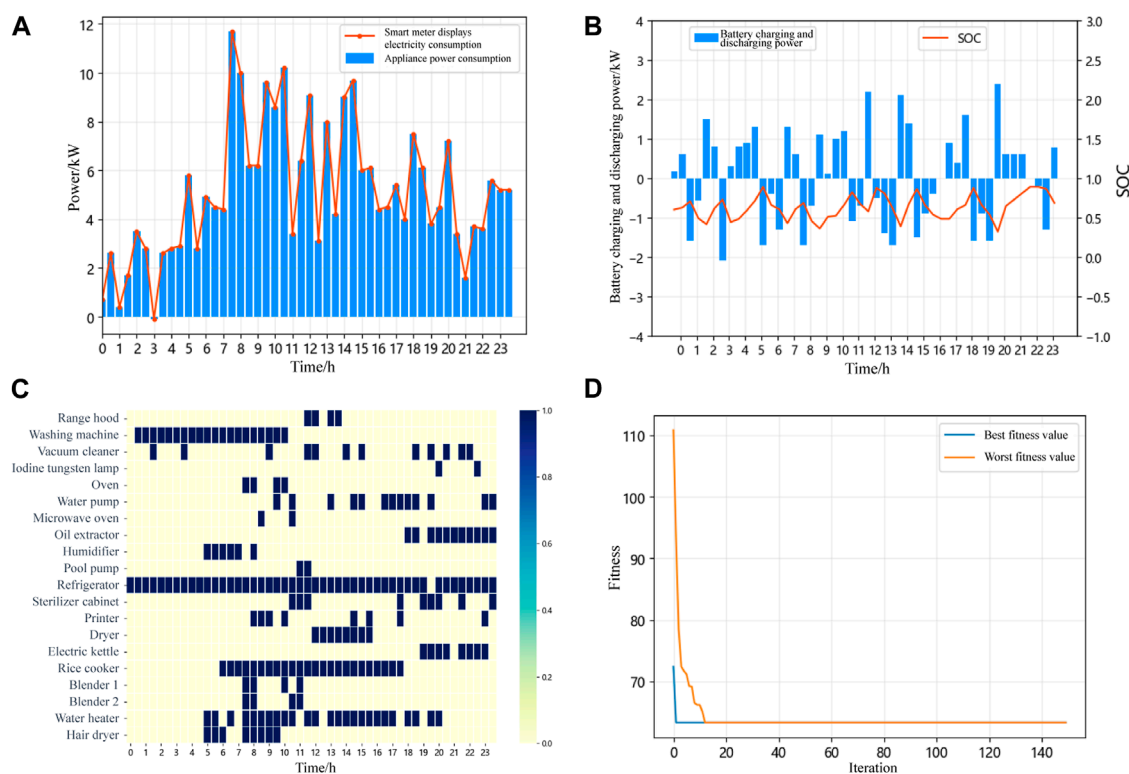


FIGURE 8

Simulation results with battery A integration: (A) electricity task scheduling results; (B) battery charging and discharging power and SOC; (C) operational status diagram of electrical devices; (D) convergence curve of genetic algorithm.

portion of this time falls within the higher electricity price period, such as 6:00–9:30. Therefore, the user's electricity cost is relatively high.

Figure 6B displays the electricity scheduling results for various adjustable devices over the 48 time slots of a day through a heatmap. Taking the vacuum cleaner as an example, its usage time reaches 2.5 h during the high electricity price periods of 5:30–9:30 and 14:30–20:30.

1) Scenario I: No battery storage is integrated into the system.

The electricity task scheduling simulation results without battery storage integration are shown in Figure 7B. The household electricity cost is 72.716 cents. Since there is no energy storage device connected, the amount of electricity exchanged with the grid for each time slot equals the consumption value from the smart meter. The figure shows that the peak electricity consumption periods are concentrated around 7:30–10:00, 12:00–12:30, and 13:30–15:00. The real-time electricity price plays a dominant role, and to save costs, users shift their electricity consumption to periods with lower prices, such as 9:30–10:00, 12:00–12:30, and 13:30–14:30. Figure 7A illustrates the indoor temperature variation curve. As shown, the indoor temperature remains within the user-defined range of 24°C–26°C.

Figure 7C displays the electricity scheduling results of various dispatchable devices over the 48 time slots in a day using a heatmap. As illustrated by the chart, the washing machine and rice cooker,

being non-interruptible devices, must complete their respective tasks once they start before they can stop. The working hours of the dispatchable devices have been adjusted accordingly. For instance, the vacuum cleaner's usage during the high electricity price periods has been reduced to 1.5 h.

Figure 7D displays the convergence curve of the genetic algorithm; the improved genetic algorithm converges relatively quickly.

2) Scenario II: System connected to Battery A or B

(1) Connection to Battery A

With the system integrated with Battery A, the electricity cost is 57.572 cents. The load scheduling simulation result is shown in Figure 8A. With the integration of a battery, the user can sell surplus electricity back to the grid. Thus, the system's feed-in capability can enhance the overall economic benefits of the system. At this point, the smart meter's displayed electricity consumption includes the electricity consumption of basic appliances and the rechargeable battery. The chart shows that the user sells electricity to the grid between 3:00 and 3:30. The electricity scheduling result now depends not only on real-time electricity prices but also on the battery's charging and discharging behavior. When the battery is charging, the power is positive, and when discharging, the power is negative. Figure 8B reveals that through energy storage with the battery, it charges during low-price phases such as 0:00–0:30, 1:30–2:30, 3:00–5:00, 9:30–11:00, 14:00–14:30, 20:30–21:30, and 23:00–24:00.

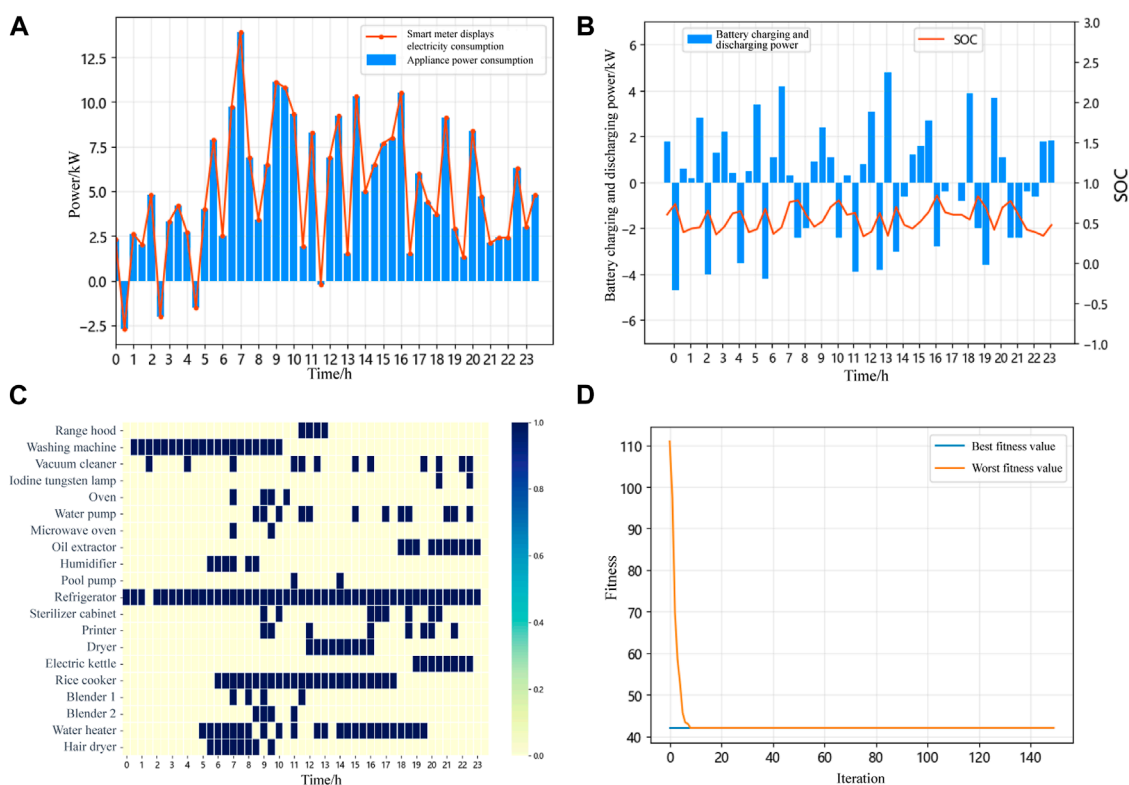


FIGURE 9

Simulation results with battery B integration: (A) electricity task scheduling results; (B) battery charging and discharging power and SOC; (C) operational status diagram of electrical devices; (D) convergence curve of genetic algorithm.

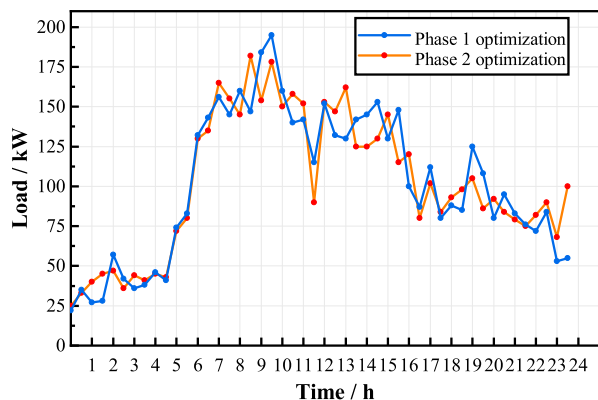


FIGURE 10

Optimization of system load curve at different stages.

During high-price periods, it discharges energy for appliances, such as 5:30–6:30, 7:30–8:30, 14:30–16:00, and 18:00–19:30. The battery's SOC value always remains between 0.3 and 0.9.

Figure 8C displays a heatmap showcasing the electricity scheduling results for various adjustable devices over a 24-h period with 48 time slots when the system is connected to a battery. There have been certain adjustments in the electricity usage of

the appliances. Taking the vacuum cleaner as an example, its usage during the high electricity price intervals is reduced to 2 h. Figure 8D illustrates the convergence curve of the genetic algorithm.

(2) Connection to Battery B

The electricity cost now stands at 42.064 cents. This implies that the larger the capacity of the battery integrated into the system, the more it aids in leveling the peaks and troughs through energy storage, resulting in a lower electricity cost for the user. The energy scheduling simulation results are depicted in Figure 9A. The task allocation for electricity devices does not differ significantly from that when Battery A was incorporated, and there is not a significant difference in the frequency of battery charge-discharge cycles. Due to the larger capacity of Battery B, excess energy is sold back to the grid during the low-price intervals, such as 00:30–1:00, 2:30–3:00, 4:30–5:00, and 11:30–12:00, leading to even lower electricity costs. However, load scheduling based on battery storage resulted in a higher peak power demand. In the figure, the power demand peak is 7:00–7:30, and this peak value is even higher than the initial peak demand of the user, posing challenges for stable operation of the system. Therefore, subsequent optimization is required in the second phase to reduce system load fluctuations and prevent new electricity demand peaks.

From Figure 9B, it can be observed that the battery is used for energy storage and is charged during the low-price intervals, such as 0:30–2:00, 2:30–4:00, 4:30–5:30, 9:30–10:00, 11:30–12:30,

TABLE 2 System load parameter optimization at different stages for the user group.

Optimization stage	Load Peak/kW	Load Valley/kW	Absolute peak-valley difference	Peak-valley coefficient	Peak-valley rate	Load factor (%)	Load fluctuation variance
First Stage Optimization	195.4	23.4	172.0	8.35	0.88	51.7	2071.6
Second Stage Optimization	182.0	25.1	156.9	7.25	0.86	56.2	1909.1

TABLE 3 Cost optimization for users at various stages.

User number	Initial cost/cents	First phase optimization cost/cents	Distributed optimization cost/cents	User number	Initial cost/Cents	First phase optimization cost/cents	Distributed optimization cost/Cents
1	119.163	60.902	56.732	11	111.973	57.568	63.593
2	117.366	58.425	66.104	12	110.496	58.742	77.725
3	105.606	64.075	51.527	13	104.334	62.916	72.560
4	110.654	61.150	80.581	14	109.976	58.137	69.696
5	116.923	58.220	63.762	15	120.054	62.180	59.515
6	118.215	61.038	60.143	16	111.585	56.630	57.222
7	121.014	59.715	56.525	17	111.355	61.610	66.568
8	108.976	55.953	68.141	18	113.335	64.514	64.415
9	114.595	58.906	80.490	19	109.225	65.378	56.531
10	113.464	58.962	69.918	20	130.775	64.775	63.685

13:00–13:30, 20:30–21:30, and 22:30–24:00, primarily in the early hours of the morning. During high-price intervals, such as 5:30–6:00, 7:30–8:30, 16:00–17:00, 17:30–18:00, and 18:30–19:30, the stored energy in the battery is released to power the devices. Similarly, the SOC of the battery consistently remains between 0.3 and 0.9.

Figure 9C displays a heatmap illustrating the electricity scheduling results of various schedulable devices over 24 h, divided into 48 time slots, when Battery B is integrated into the system. Adjustments can be observed in the usage times of devices such as the vacuum cleaner and the disinfection cabinet. Taking the vacuum cleaner as an example, its initial load usage during high electricity price periods was 2.5 h, which has now been reduced to 2 h. Figure 9D presents the convergence curve of the genetic algorithm.

6.3 Second-stage simulation results

Demand response can reduce user costs, but it might introduce new peak electricity demands to the system. Therefore, a second phase of distributed optimization is conducted to minimize system load fluctuations. Considering an area with $W = 20$ households, a distributed load fluctuation optimization simulation experiment

is conducted. The energy scheduling mechanism's impact on user electricity consumption behavior is analyzed from the perspective of a group of users. The energy storage system opted to integrate Battery A. The system load command is set at $\min \text{VAR}(P_h^{\text{Total}}) \leq 0.95 * \text{VAR}(P_{h,0}^{\text{Total}})$, which means optimization stops when this condition is met. In this context, $\text{VAR}(P_{h,0}^{\text{Total}})$ represents the load fluctuation level from the user's self-optimization in the first phase.

Figure 10 shows the results of the system load fluctuation optimization simulation. As evident from the figure, after undergoing system load fluctuation optimization, compared to the self-optimization of user costs in the first phase, there is a significant change in the electricity consumption patterns of the user group within the area. The system load curve becomes more stable, with a reduction in peak values and an increase in valley values.

Table 2 shows the electricity consumption behavior characteristics of the user group. The quality of system electricity consumption is assessed through indicators such as the load factor, which is the ratio of the average load to the peak load over a specific period. Improving the load factor can effectively reduce peaks, elevate valleys, decrease the peak-valley difference, and ensure the safe and stable operation of the power system. From Table 2, it can be observed that after the second-phase load fluctuation optimization,

the system's peak load decreased by 6.9%, the valley increased by 7.2%, the absolute peak-valley difference dropped by 8.8%, the peak-valley coefficient decreased by 13.2%, and the peak-valley difference rate reduced by 2.3%. Additionally, the load factor rose from 51.7% to 56.2%, an increase of 4.5%, and the variance of the load fluctuation decreased by 7.8%.

Table 3 shows the user costs before and after distributed optimization. The average cost for users before and after distributed optimization decreased from 113.954 cents to 65.272 cents, a reduction of 42.7%. This shows that user costs decrease after distributed optimization. Compared to that in the first stage of optimization, the costs for some users increased because they made financial sacrifices to change system load fluctuations. Users can set the weighting factor γ as needed to balance cost and load fluctuation. Considering real-world scenarios, power companies can incentivize users who have increased costs, for instance, by reducing electricity prices, to encourage them to participate in optimizing load fluctuations.

7 Conclusion

This paper proposed a two-stage distributed optimization method for the HEMS based on data-driven algorithm. Firstly, a distributed load scheduling framework for HEMS is established, and various devices are modeled. Secondly, a two-stage optimization method is introduced, targeting both the minimization of user cost and load fluctuations to achieve demand response. Finally, simulation experiments of the load scheduling are conducted, and the impact of battery parameters on energy scheduling is analyzed. Simulation results demonstrate that users can adjust their loads based on comfort and the urgency of device usage. The main conclusions are as follows:

- The first optimization stage results indicate that when the battery capacity integrated into the system increases from 3.68 kWh to 6.68 kWh, user costs can be reduced from 57.572 cents to 42.064 cents. It is evident that not only can the proposed method effectively save electricity costs for users, but the introduction of larger capacity batteries also significantly reduces these costs.
- The second stage results indicate that, the system's peak load decreases by 6.9%, the valley increases by 7.2%, and the absolute peak-valley difference is reduced by 8.8%. This demonstrates

that the proposed method can effectively optimize the usage data of a group of users and decrease system load fluctuations.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

FC: Conceptualization, Investigation, Validation, Writing–original draft. DA: Funding acquisition, Project administration, Writing–review and editing. YZ: Investigation, Validation, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China under Grant 61803295; in part by the Major Research Plan of the National Natural Science Foundation of China under Grant 61833015.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Awad, A., Bazan, P., and German, R. (2015). "Privacy aware demand response and smart metering," in 2015 IEEE 81st Vehicular Technology Conference (VTC Spring) (IEEE), Glasgow, UK, 11–14 May 2015. 1–5.
- Bejoy, E., Islam, S., and Oo, A. (2017). "Optimal scheduling of appliances through residential energy management," in 2017 Australasian Universities Power Engineering Conference (AUPEC) (IEEE), Genre: Konferenzschrift, 19–22 Nov. 2017.
- Chen, S.-Y., and Chang, C.-H. (2023). Optimal power flows control for home energy management with renewable energy and energy storage systems. *IEEE Trans. Energy Convers.* 38, 218–229. doi:10.1109/tec.2022.3198883
- Duman, A. C., Erden, H. S., Gönül, Ö., and Güler, Ö. (2021). A home energy management system with an integrated smart thermostat for demand response in smart grids. *Sustain. Cities Soc.* 65, 102639. doi:10.1016/j.scs.2020.102639
- Fekri, M. N., Patel, H., Grolinger, K., and Sharma, V. (2021). Deep learning for load forecasting with smart meter data: online adaptive recurrent neural network. *Appl. Energy* 282, 116177. doi:10.1016/j.apenergy.2020.116177
- Huang, C., Wang, X., Gan, Q., Huang, D., Yao, M., and Lin, Y. (2021). A lightweight and fault-tolerable data aggregation scheme for privacy-friendly smart grids environment. *Clust. Comput.* 24, 3495–3514. doi:10.1007/s10586-021-03345-w
- Jiang, X., and Wu, L. (2020). Residential power scheduling based on cost efficiency for demand response in smart grid. *IEEE Access* 8, 197324–197336. doi:10.1109/access.2020.3034767
- Joo, I.-Y., and Choi, D.-H. (2017). Distributed optimization framework for energy management of multiple smart homes with distributed energy resources. *IEEE Access* 5, 15551–15560. doi:10.1109/access.2017.2734911

- Kou, X., Li, F., Dong, J., Starke, M., Munk, J., Kuruganti, T., et al. (2019). "A distributed energy management approach for residential demand response," in 2019 3rd International Conference on Smart Grid and Smart Cities (ICSGSC) (IEEE), Berkeley, CA, USA, June 25 2019 to June 28 2019, 170–175.
- Li, C., Chen, Y., Yang, Y., Li, C., and Zeng, Y. (2019). "Ppcsb: a privacy-preserving electricity consumption statistics and billing scheme in smart grid," in Artificial Intelligence and Security: 5th International Conference, ICAIS 2019, New York, NY, USA, July 26–28, 2019, 529–541.
- Li, J., Guo, L., Zuo, Y., and Liu, W. (2022). A design method for wideband chaff element using simulated annealing algorithm. *IEEE Antennas Wirel. Propag. Lett.* 21, 1208–1212. doi:10.1109/lawp.2022.3161762
- Lin, Y.-H. (2021). Trainingless multi-objective evolutionary computing-based nonintrusive load monitoring: Part of smart-home energy management for demand-side management. *J. Build. Eng.* 33, 101601. doi:10.1016/j.jobbe.2020.101601
- Lu, R., Heung, K., Lashkari, A. H., and Ghorbani, A. A. (2017). A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced iot. *IEEE access* 5, 3302–3312. doi:10.1109/access.2017.2677520
- Mahela, O. P., Khosravy, M., Gupta, N., Khan, B., Alhelou, H. H., Mahla, R., et al. (2022). Comprehensive overview of multi-agent systems for controlling smart grids. *CSEE J. Power Energy Syst.* 8, 115–131. doi:10.17775/CSEEJPES.2020.03390
- Ming, Z., Geng, W., Haojing, W., Ran, L., Bo, Z., and Chenjun, S. (2016). Regulation strategies of demand response considering user satisfaction under smart power background. *Power Syst. Technol.* 40, 2917–2923. doi:10.13335/j.1000-3673.pst.2016.10.001
- Munankarmi, P., Wu, H., Pratt, A., Lunacek, M., Balamurugan, S. P., and Spitsen, P. (2022). Home energy management system for price-responsive operation of consumer technologies under an export rate. *IEEE Access* 10, 50087–50099. doi:10.1109/access.2022.3172696
- Pal, S., Kumar, M., and Kumar, R. (2017). "Price aware residential demand response with renewable sources and electric vehicle," in 2017 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE) (IEEE), Dehradun, India, 18–19 December 2017, 211. –214.
- Sangswang, A., and Konghirun, M. (2020). Optimal strategies in home energy management system integrating solar power, energy storage, and vehicle-to-grid for grid support and energy efficiency. *IEEE Trans. Industry Appl.* 56, 5716–5728. doi:10.1109/tia.2020.2991652
- Seal, S., Boulet, B., Dehkordi, V. R., Bouffard, F., and Joos, G. (2023). Centralized mpc for home energy management with ev as mobile energy storage unit. *IEEE Trans. Sustain. Energy* 14, 1425–1435. doi:10.1109/tste.2023.3235703
- Wang, Z., Munawar, U., and Paranjape, R. (2020). "Stochastic optimization for residential demand response under time of use," in 2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020) (IEEE), Le Méridien Cochin, Kerala, India, 2–4 Jan 2020.
- Wei, H., Bo, Z., and Yubo, L. (2019). eal-time pricing scheme based on privacy protection. *Appl. Res. Computers/Jisuanji Yingyong Yanjiu* 36. doi:10.19734/j.issn.1001-3695.2017.12.0823
- Yang, J., Huang, G., and Wei, C. (2018). Privacy-aware electricity scheduling for home energy management system. *Peer-to-Peer Netw. Appl.* 11, 309–317. doi:10.1007/s12083-016-0492-x
- Yang, Z., Pu, Y., Liu, H., and Ni, M. (2020). "Optimal home energy management strategy considering integrated demand response," in 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Nanjing, China, 20–23 September 2020.
- Zhao, Q., and Li, C. (2020). Two-stage multi-swarm particle swarm optimizer for unconstrained and constrained global optimization. *IEEE Access* 8, 124905–124927. doi:10.1109/access.2020.3007743



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Kaixun He,
Shandong University of Science and
Technology, China
Longchao Yao,
Zhejiang University, China
Guodong Yang,
Chinese Academy of Sciences (CAS),
China

*CORRESPONDENCE

Siyu Miao,
✉ siyumiao_hbu@163.com

RECEIVED 01 September 2023

ACCEPTED 23 October 2023

PUBLISHED 03 November 2023

CITATION

Zhao Z, Lv X, Xi Y and Miao S (2023),
Defect detection method for key area
guided transmission line components
based on knowledge distillation.
Front. Energy Res. 11:1287024.
doi: 10.3389/fenrg.2023.1287024

COPYRIGHT

© 2023 Zhao, Lv, Xi and Miao. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Defect detection method for key area guided transmission line components based on knowledge distillation

Zhenbing Zhao^{1,2,3}, Xuechun Lv^{1,4}, Yue Xi¹ and Siyu Miao^{1*}

¹School of Electrical and Electronic Engineering, North China Electric Power University, Baoding, China,

²Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education, North China Electric Power University, Baoding, China, ³Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding, China, ⁴Institute of Electrical Automation, China Nuclear Power Engineering Co., Ltd., Hebei Branch, Shijiazhuang, China

Introduction: The aim of this paper is to address the problem of the limited number of defect images for both metal tools and insulators, as well as the small range of defect features.

Methods: A defect detection method for key area-guided transmission line components based on knowledge distillation is proposed. First, the PGW (Prediction-Guided Weighting) module is introduced to improve the foreground target distillation region, and the distillation range is precisely concentrated in the position of the first k feature pixels with the highest quality score in the form of a mask. The feature knowledge of defects of hardware and insulators is used as the focus for the teacher network to guide the student network. Then, the GcBlock module is used to capture the relationship between the target defects of the hardware and the transmission lines in the background, and the overall relationship information of the image is used to promote the students' network to learn the teacher's network perception ability of the relationship information. Finally, the classification task mask and regression task mask generated by the PGW module, combined with the overall image relationship loss, form a distillation loss function for network training to improve the accuracy of students' network detection accuracy.

Results and Discussion: The effectiveness of the proposed method is verified by using self-build metal fittings and insulator defect data sets. The experimental results show that the student network mAP₅₀ (Mean Average Precision at 50) in the Faster R-CNN model with the knowledge distillation algorithm added in this paper increases by 8.44%, and the RetinaNet model increases by 2.6%. The Cascade R-CNN model improved by 5.28%.

KEYWORDS

knowledge distillation, key region guidance, component defects, teacher model, student model

1 Introduction

The transmission line is one of the most important infrastructures of China's energy Internet, and ensuring the reliability of the transmission line is one of the important contents of the construction of the energy Internet. Transmission line components are an important part of mechanical connection, fixing, protection and insulation. However, they are

susceptible to abnormalities and defects such as defects, corrosion and soiling due to the influence of the complex natural environment and harsh climatic conditions (Nguyen et al., 2018). Therefore, regular inspection and maintenance of transmission line fittings, insulators and other important components can effectively reduce a series of safety accidents caused by transmission line faults.

Inspection is a way to guarantee the continuous and stable power supply of the power grid, the purpose of which is to carry out online condition detection and fault diagnosis of components such as shockproof hammers and insulators on the lines (Zhao and Cui, 2018). The current transmission line inspection methods include manual inspection, robot inspection (Toth and Gilpin-Jackson, 2010), helicopter (Pham et al., 2020), unmanned aerial vehicle inspection (Li et al., 2021), and remote sensing satellite inspection (Yang et al., 2021). "Drone inspection is the main focus, supplemented by manual labor" has developed into the main operation and maintenance mode of China's power system (Yang et al., 2020). The construction of intelligent and manual synergistic inspection system is an important initiative to promote the safe operation of the power grid (Du et al., 2022).

The development of deep learning technology provides an effective means for transmission line inspection and can more effectively complete the task of defect detection of transmission components in aerial images. At present, a lot of research work has been done. Literature (Zhai et al., 2023) proposes a transmission line multi-fitting detection method based on implicit spatial knowledge fusion, aiming at the tiny-size and dense occlusion problem in the transmission line multi-fitting detection task. First, in order to mine the implicit spatial knowledge between transmission line fittings to assist the model in detection, the spatial box setting module and the spatial context extraction module are proposed to set the spatial box and extract the spatial context information. Then, the spatial context memory module is designed to filter and remember the spatial context information to assist the location of the multi-fitting detection model. Finally, the post-processing part of the model is improved to further alleviate the low detection accuracy problem caused by dense occlusion fittings. The experimental results show that the proposed model has a promotion effect on the detection of various kinds of fitting. Literature (Li et al., 2023) proposes a metal fittings equipment detection algorithm based on improved YOLOV7. This method adds a CA attention mechanism to the network structure of YOLOV7 to enhance the feature extraction of hardware devices in the network model. At the same time, it reduces the interference of complex backgrounds on the network model to extract features of hardware devices, allowing the network model to extract features in detail, thereby improving the network model's detection generalization for hardware devices. In order to alleviate the problem of misdetection and recheck caused by the lack of context information in various existing hardware and defect detection methods, literature (Zhao et al., 2023) proposes a method of transmission line hardware and defect detection based on context-structure reasoning. First, the image is input into the target detection model; Then, the output result of the detection model is sent to the structural reasoning module, and the output result is sent to the bidirectional gated cycle unit and self-attention for processing. The structural knowledge of transmission line fittings and their defects is used to improve the confidence degree of the correct positive sample and reduce the confidence degree of the

wrong positive sample. Finally, the final output result is obtained through the regressor. To achieve the purpose of improving the average accuracy. Literature (Sun et al., 2023) proposes a two-stage insulator defect detection framework composed of attention-based insulator detection network and defect detection network. Among them, the attention-based insulator detection network is responsible for the location of the insulator, and the defect detection model determines whether the insulator is damaged. The two-stage design of first positioning and then detection avoids the interference of complex background and can realize the high-precision detection of defects. Literature (Li et al., 2023) proposes a multi-scale feature fusion insulator defect detection network for solving the problem of insulator defective regions with little pixel information and varying shapes and sizes. The network used a residual attention network to obtain insulator defect features with different resolutions, and designed a multi-scale feature fusion network based on inverse convolution and multi-branch detection, which gradually fused the deep feature maps with the shallow feature maps. In this way, more abundant image semantic information can be generated for target classification and location regression. In addition, the literature also used Focal loss and Gaussian non-great suppression methods to further enhance the detection effect.

Although these methods improve the accuracy of detection, they inevitably increase the complexity of the model, consume a lot of computing resources and time, and are difficult to deploy on resource-limited equipment. The knowledge distillation algorithm provides a solution to this problem. Literature (Gu et al., 2023) proposes a deep neural network model compression algorithm for knowledge distillation of multi-teacher models, which takes advantage of the integration of multi-teacher models and takes the predictive cross-entropy of each teacher model as the quantitative criterion for screening to select the teacher model with better performance to guide students, and allows the student model to extract information from the feature layer of the teacher model. And give better performing teacher models a greater say in instruction. Literature (Wang et al., 2022) proposes an attention mechanism based on the feature map quality evaluation algorithm (IQE). The knowledge distillation method based on the IQE attention mechanism uses the IQE method to identify important knowledge in the pre-trained SAR target recognition deep neural network. Then in the process of knowledge distillation, the lightweight network is forced to focus on the learning of important knowledge. Through this mechanism, the method proposed in this paper can efficiently transfer the knowledge of the pre-trained SAR target recognition network to the lightweight network, which makes it possible to deploy the SAR target recognition algorithm on the edge computing platform. Literature (Zhao et al., 2022) propose a target detection model distillation (TDMD) framework using feature transition and label registration for remote sensing imagery. A lightweight attention network is designed by ranking the importance of the convolutional feature layers in the teacher network. Multiscale feature transition based on a feature pyramid is utilized to constrain the feature maps of the student network. A label registration procedure is proposed to improve the TDMD model's learning ability of the output distribution of the teacher network.

At present, some researches have applied the knowledge distillation method to the field of electric power. Literature (Yang et al., 2022) proposes a compression and integration application

method based on knowledge distillation. In this method, the Detr model is used to identify the initial target, and the Deformable Detr algorithm is used to compress the Detr model, so that the compression ratio reaches 87.5% and the target detection accuracy is maintained at a high level, and the effective integrated application of the target detection model in the substation inspection robot body is realized. Literature (Zhao et al., 2021) proposes a bolt defect image classification method based on dynamically supervised knowledge distillation, aiming to solve the problem of high computational resource consumption of large models. The method utilizes adaptive weighting and attention transfer techniques to improve the ability of the small model to learn and represent bolt defects, which in turn enhances its classification performance. In addition, literature (Zhang et al., 2022) improved the YOLOv4 model and introduced the PCSA (Positional Contextual Attention Shift) attention module for the problem of anti-vibration hammer small target detection. This method combines pruning and knowledge distillation techniques to tailor and compress the network parameters, and constructs a lightweight anti-vibration hammer detection network model, PCSA-YOLOs, to improve the detection accuracy of small targets in complex backgrounds.

With the intelligent development of power system inspection technology, it is urgent to deploy models on UAV and helicopter aerial photography and online monitoring equipment. Knowledge distillation can help improve the performance of the model with a small number of parameters, but making the student network simulate the teacher's network feature extraction ability without difference cannot achieve the best effect. The focus of this paper is to enable students to learn the effective feature processing ability of teachers' networks. In this paper, the PGW module is first introduced to refine the feature knowledge of foreground object distillation, and the first k most important pixels are extracted to form a feature mask to improve the distillation performance of student network for detecting hardware defects. Then, the GcBlock module is used to capture the relationship between the target defects of the hardware and the transmission lines in the background, and the overall relationship information of the image is used to promote the students' network to learn the teacher's network perception ability of the relationship information. The combination captures information about the relationship between transmission line components and components, and between components and backgrounds, helping to improve target detection accuracy.

2 Research methodology

2.1 Refinement of prospects regional distillation

The application of knowledge distillation algorithms to image target detection has focused on models using feature pyramid networks. Past approaches usually directly used the output of the classification and regression tasks of the teacher network as the target of the student network. With the development, knowledge distillation can guide the training of the student network in a more targeted way to improve the detection accuracy. As shown in the literature (Guo et al., 2021), unlike the general classification task, the classification and regression tasks in

the detection network can be negatively affected if the same objective function is used for both tasks. This is because the two tasks have different preferences for features: classification requires regions with rich semantic information, whereas regression prefers to focus on the edge portion of feature information. Features that produce better classification scores are not accurate enough in predicting bounding boxes (Song et al., 2020). Therefore, the same sensory field does not guarantee optimal performance for both classification and regression tasks. As shown by the images of defects of gold tools and insulators, important feature knowledge exists for defects of the same kinds of gold tools and insulators, and the datasets of defects of gold tools and insulators are much smaller, which should be fully utilized to guide the students' network by taking advantage of the ability of the teacher's network to deal with the feature knowledge of the defects.

Therefore the features are scored to determine whether they are good for classification or regression tasks, in order to reduce the adverse effect of complex background on the detection of defective targets of gildings and insulators, with the help of ground-truth box first decouple each layer of features of FPN (Feature Pyramid Network) whether it belongs to the foreground target region or the background region as shown in Eq. 1:

$$M_{(i,j)} = \begin{cases} 1, & \text{if } (i,j) \in G \\ 0, & \text{if } (i,j) \notin G \end{cases} \quad (1)$$

Let (i,j) denote the horizontal and vertical coordinates of the feature pyramid network generating the feature map of the model. If (i,j) is in the ground-truth bounding box, it is determined that this feature belongs to the foreground target region, and the mask is set $M_{(i,j)} = 1$; if (i,j) is not in the ground-truth bounding box, it is determined to be the feature map of the background region, and the mask is set $M_{(i,j)} = 0$.

Amplifying the most meaningful feature distillation signals generated by the teacher network and using them to guide the student network is the purpose of knowledge distillation. For this purpose, we look at the quality of a teacher's bounding box predictions taking both classification and localization into consideration. Formally, the quality score of a box $b_{(i,j)}$ predicted from a position $X_i = (x_i, y_i)$ w.r.t. a ground truth b is as shown in formula (2):

$$q(b_{(i,j)}, G) = M_{(i,j)} \cdot (p_{(i,j)} G)^\lambda \cdot (\text{DIOU}(G, b_{(i,j)}))^{1-\lambda} \quad (2)$$

where $M_{(i,j)}$ is an indicator function that is 1 if X_i lies inside box b and 0 otherwise; $(p_{(i,j)} G)$ is the classification probability w.r.t. the GT box's category; $\text{DIOU}(G, b_{(i,j)})$ is the DIOU between the predicted and ground-truth box. $b_{(i,j)}$ is a prediction frame and G is a real labeled box, it is the ground-truth box; λ is a hyper-parameter that balances classification and localisation. We calculate the quality score of location X_i as the maximum value of all prediction scores for that particular location, as shown in Eq. 3. Y is used to represent the set of X_i locations as shown in Eq. 4:

$$q_i = \max_{j \in Y} q(b_{(i,j)}, G) \quad (3)$$

$$Y = \{1, 2, \dots, X_i\} \quad (4)$$

Use of mass fraction q_i as an important parameter in determining distillation, these positions are the highest quality predictive ensemble of scores generated by the teacher network.

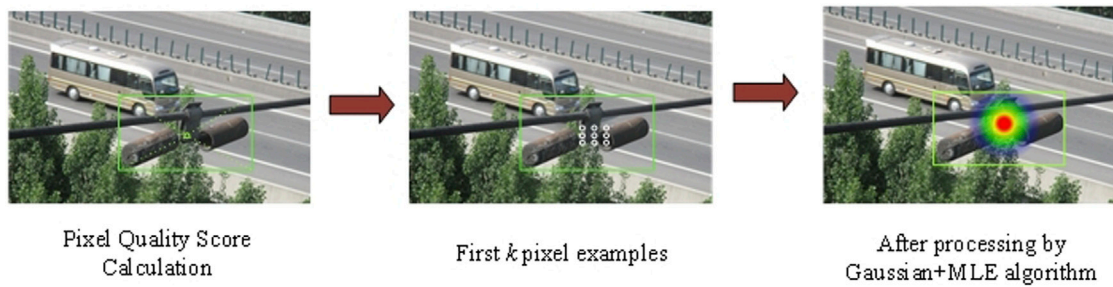


FIGURE 1
Schematic diagram of the processing principle for the foreground target area.

The purpose of knowledge distillation is to allow the student network to mimic the strong generalization ability of the teacher network. To achieve this, this paper focuses the foreground distillation region on the locations where the teacher network produces predictions with high-quality scores, as these locations contain detection information that combines both classification and regression considerations and represent the excellent performance of the teacher network, helping the student network to improve its ability to detect defective targets in complex contexts.

2.2 Knowledge distillation guided by key area scoring

The features of fitting and insulator defects are very different from those of their intact targets; the features are more consistent between parts of the same species, but the features of each component defect are variable and complex (Zhao et al., 2021), and when labeling the dataset, the size of the labeled box is as close to the target as possible, and the defects are damages that are produced on the component targets, which have a more reduced range of effective features compared to the intact targets. Therefore, we would like to use the knowledge of features of defects in gold tools and insulators as a focus for the instructor network to guide the student network. Therefore, the PGW (Prediction-Guided Weighting) (Yang et al., 2022) module is introduced to improve the prospect distillation region. And the PGW module is precisely concentrated in the first k feature pixels with the highest mass fraction in the prospect region. The effect of each position is then smoothed according to the two-dimensional Gaussian distribution fitted for each ground-truth box by the maximum likelihood estimation method. Finally, only the k position is extracted in the foreground target region, and the weight of the position is assigned by the Gaussian function, the schematic is shown in Figure 1.

We smooth the effects of each position according to the 2D Gaussian distribution fitted by the maximum likelihood estimate (MLE) for each ground-truth box. Finally, foreground distillation is performed only for those k positions, whose weights are assigned by Gaussian. For detection targets with ground-truth O , the quality score q_i for each feature pixel within G is first calculated (Du et al., 2021). The calculation formula of q_i is shown in (3). The k highest scoring pixels are then selected among all the layers of the FPN network. $T^O = \{(X_k^O, L_k^O) | k = 1, \dots, K\}$ is used to generalize the k highest scoring pixels. Where X_k^O is the absolute coordinate of the

pixel of the detection target O . L_k^O denotes which layer of the FPN where the pixel of the detection target O is located. Both quantities represent the k th pixel. Assuming that the selected pixel description is represented on the image plane as $T_k^O \sim N(\mu, \Sigma | O)$ defined, the maximum likelihood estimation algorithm is used to compute and the two parameters μ and Σ as shown in Eqs 5, 6:

$$\mu = \frac{1}{K} \sum_{k=1}^K X_k^O \quad (5)$$

$$\Sigma = \frac{1}{K} \sum_{k=1}^K (X_k^O - \mu)(X_k^O - \mu)^T \quad (6)$$

Each feature pixel $p_{(i,j)l}$ with absolute $X_{(i,j)}$ coordinates in layer l of the FPN, calculates its importance in the distillation process as expressed in Eq. 7:

$$I_{(i,j)l}^O = \begin{cases} 0 & P_{(i,j)l} \notin T^O \\ \exp\left(\left(-\frac{1}{2}(X_{(i,j)} - \mu)\Sigma^{-1}(X_{(i,j)} - \mu)^T\right)\right) & P_{(i,j)l} \in T^O \end{cases} \quad (7)$$

where $I_{(i,j)l}^O$ denotes that the detection target O is in layer l of the FPN, the importance of the feature pixel $p_{(i,j)l}$ with coordinates (i, j) . If $p_{(i,j)l}$ belongs to the highest scoring pixel, calculate the importance value using the formula, and if it does not belong to the pixel, make it equal to zero. If a feature pixel is equally important for more than one object, we use its maximum value $I_{(i,j)l}$, and the formula representation is shown in (8):

$$I_{(i,j)l} = \max_o \{I_{(i,j)l}^O\} \quad (8)$$

l refers to a layer of the FPN layer of size $H_l \times W_l$, by normalizing the importance of the distillation with the number of pixels of that layer that are important and have a non-zero mass fraction so that they are used to assign the distillation weights Q , and the formula is expressed as shown in (9).

$$Q = \frac{I_{(i,j)l}}{\sum_{i=1}^{H_l} \sum_{j=1}^{W_l} M_{(i,j)l}} \quad (9)$$

H_l and W_l denote the length and width dimensions of the feature map in the l -layer of the FPN, the above process constitutes the Predictive Guidance Weighting (PGW) module, whose output is the

foreground distillation weight Q of all feature levels and pixels, which are used to find out the important feature pixels for use in the teacher's network to guide the student's network.

2.3 Overall distillation loss function

The high-level network of the feature extraction network can output better semantic features, but due to the size of the feature map is too small, the geometric information is not sufficient, which is not conducive to the detection of the target; the shallow network contains more geometric information, but the semantic features of the image are not much, which is not conducive to the classification of the image. Therefore, when the network is trained, the student network is made to learn the ability of the teacher network to extract and process features at each layer of the FPN. By distilling the classification and regression tasks separately, the student network integrally learns the generalization ability of the teacher network for these two tasks, which leads to an increase in its detection accuracy.

Formally, at each feature level in FPN, this paper utilizes the PGW module to generate two distinct foreground distillation masks, $Q_{(i,j)}^{cls}$ is the foreground distillation mask generated specifically for the classification task. $Q_{(i,j)}^{reg}$ is the foreground distillation mask generated specifically for the regression task. In this context, $R^{C \times H \times W}$ represents a feature layer has C channels, and the feature map height and width of each layer H and W . During training, the student model is encouraged to learn the corresponding classification and regression features $F_{n(i,j)}^{S,cls}, F_{n(i,j)}^{S,reg} \in R^{C \times H \times W}$ from the teacher network. The classification feature loss function is represented as Eq. 10, and the regression feature loss function is represented as Eq. 11.

$$L_{fea}^{cls} = \sum_{n=1}^C \sum_{i=1}^H \sum_{j=1}^W \left(\alpha Q_{(i,j)}^{cls} + \beta N_{(i,j)}^{cls} \right) \left(F_{n(i,j)}^{T,cls} - F_{n(i,j)}^{S,cls} \right)^2 \quad (10)$$

$$L_{fea}^{reg} = \sum_{n=1}^C \sum_{i=1}^H \sum_{j=1}^W \gamma Q_{(i,j)}^{reg} \left(F_{n(i,j)}^{T,reg} - F_{n(i,j)}^{S,reg} \right)^2 \quad (11)$$

α , β , and γ are hyperparameters used to balance the loss weights. $N_{(i,j)}^{cls}$ is the normalized mask for the background distillation region. As shown in Eq. 12, when $Q_{(i,j)}^{reg} \neq 0$, $M_{(i,j)}$ is 1. The meaning of $\mathcal{M}_{(h,w)}$ is opposite to $M_{(i,j)}$, and if the pixel (h, w) is not inside the region G , it is assigned a value of 1.

$$N_{(i,j)}^{cls} = M_{(i,j)}^{-1} / \sum_{h=1}^H \sum_{w=1}^W \mathcal{M}_{(h,w)} \quad (12)$$

In order to distill background information, the GcBlock (Cao et al., 2019) module is utilized to capture the relationship between defects in the hardware target and the transmission lines present in the background. This module leverages the overall relationship information in the image (Park et al., 2019). It encourages the student network to learn the teacher network's ability to perceive relationship information (Hu et al., 2018). The representation of the overall image relationship loss is given by Eq. 13.

$$L_{rela} = \mu \sum (R(F^T) - R(F^S))^2 \quad (13)$$

In the equation, μ is a hyperparameter used to balance the loss function, and F^T and F^S are the feature maps generated by the

teacher and student models, respectively. The function R represents the relationship information between the hardware target captured by different models and the background region. Its specific formulation is given by Eq. 14.

$$R(F_i) = F_i + W_{v2} Relu \left(LN \left(W_{v1} \left(\sum_{j=1}^{N_p} \frac{e^{W_k F_j}}{\sum_{m=1}^{N_p} e^{W_k F_m}} F_j \right) \right) \right) \quad (14)$$

In this context, W_{v1} , W_{v2} and W_k represent different linear transformation matrices. F_i represents the feature map of an input instance. $\delta_j = \frac{e^{W_k F_j}}{\sum_{m=1}^{N_p} e^{W_k F_m}}$ represents the weights obtained from global attention pooling. LN stands for Layer Normalization. N_p represents the number of positions in the feature layer. In the context of an image, $N_p = H \cdot W$. e represents the natural constant. F_M represents the feature map generated for the foreground region. F_j is any extracted feature from the image. $\delta(\cdot) = W_{v2} Relu(LN(W_{v1}(\cdot)))$ denotes the feature transformation that captures channel dependencies. The GcBlock module consists of two components: global attention pooling for context modeling and bottleneck transformation to capture channel correlations.

The structural diagram of the distillation method in this paper is shown in Figure 2. It consists of two parts: foreground object region distillation and background relationship distillation. For foreground object region distillation, the PGW module is used to calculate the masks for both the classification and regression tasks, which together form the loss function used for training the foreground object region. The background relationship distillation area adopts the GcBlock module to capture relationships in the image. The overall distillation loss function in this paper is represented as Eq. 15.

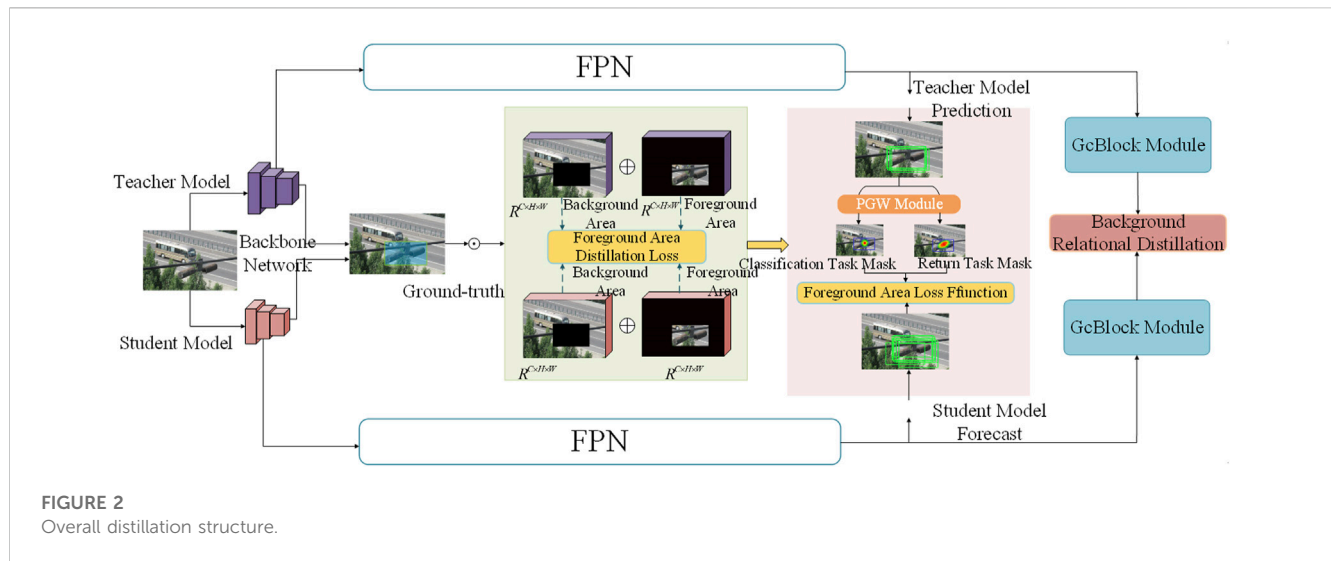
$$L = L_{fea}^{cls} + L_{fea}^{reg} + L_{rela} \quad (15)$$

3 Manuscript experimental results and analysis

3.1 Experiment preparation

This article uses 5 types of fittings and insulators, along with their defects, as the research dataset. Each type of defect has a corresponding normal target feature for comparison, including normal vibration damper, vibration damper cross, vibration damper corrosion, normal single insulator, and insulator drop. There are a total of 2,497 images, with 1997 images in the training set, 250 images in the test set, and 249 images in the evaluation set, with a ratio of 8:1:1. The dataset contains a total of 4,628 objects to be detected.

The model described in this article is trained and tested using the NVIDIA GeForce GTX 1080Ti professional accelerator card. The operating system used is Ubuntu 16.04.6 LTS, with training accelerated using CUDA 10.1. The computer language used is Python 3.7.11, and the network development framework is PyTorch. All programs are executed based on the MMDetection 2.16 toolbox. This article uses the commonly used evaluation metric



in object detection models, mean Average Precision (mAP), to assess the performance of the model. The mAP is calculated by computing the Average Precision (AP) for each class of objects and then taking the mean of all class APs. The resulting mAP is used as the final evaluation metric for the object detection model.

3.2 Comparison experiments with multiple models using distillation algorithm

In the experiments, the teacher network, student network, and the network with the distillation algorithm all use a learning rate of 0.0001. The backbone network for the teacher network is ResNet101, while for the student network, it is ResNet50. The training is conducted for 24 epochs (training rounds), and the batch size used is 50.

The experimental process in this article mainly consists of two steps.

- 1) Train the object detection models separately using the teacher network (with a larger number of parameters) and the student network (with a smaller number of parameters) on the fittings defect dataset. Calculate and record their respective accuracy.
- 2) Train the student network, which has a smaller number of parameters, using the knowledge distillation algorithm with the fittings defect dataset. After training, calculate and record its accuracy.

To validate the effectiveness of the proposed knowledge distillation algorithm, a comparative experiment is conducted using the evaluation metrics mentioned earlier. The experimental results are presented in Table 1. Three image detection models, namely, Faster R-CNN, RetinaNet, and Cascade R-CNN, are used in the experiments. Each model is separately trained as a teacher model, a student model, and a model with the knowledge distillation algorithm introduced in this article. The detailed process is to put the same data set into the same model with the backbone model of Resnet101 and Resnet50. The backbone network is identified as the teacher network by Resnet101 and the

backbone network is identified as the student network by Resnet50. The results of two different backbone network training were analyzed. After adding the distillation algorithm in this paper, the student network is trained again, and the difference between the student network with distillation algorithm added and the student network without distillation algorithm added is compared. The control parameters adopted in these processes are consistent, the learning rate is 0.0001, and the epoch of the training rounds is 24. The batch size is 50.

To verify the general applicability of the proposed method, both single-stage and two-stage models are used in the experiments. From the data in Table 1, it can be observed that the detection performance of the student network improves significantly after the knowledge distillation algorithm is applied. This distillation algorithm, as presented in this article, utilizes a decoupling approach between foreground and background information regions. By focusing on the foreground object regions and reducing the interference caused by complex backgrounds, the algorithm enhances the localization ability of the model. By setting the k value to 45 in the foreground object regions, the data indicates that, in most cases, adding the knowledge distillation algorithm proposed in this article improves the detection accuracy of the student network for both normal component targets and defects. Particularly, the improvement in detecting fittings and insulator defects is greater than the improvement in detecting normal targets. In the case of the RetinaNet model, where the teacher network's accuracy is lower than the student network's accuracy in detecting vibration damper corrosion, adding the distillation algorithm does not improve the performance of the student network. This could be because the teacher network's performance is inferior to the student network's performance, which hinders its ability to guide the student network in improving accuracy. In the case of Faster R-CNN, the highest improvement is achieved in detecting vibration damper corrosion, reaching up to 22.6%. For RetinaNet, the highest improvement is seen in detecting normal insulators, with a maximum improvement of 6.6%. Cascade R-CNN shows the highest improvement in detecting vibration damper corrosion, reaching 8.2%.

The table shown in Table 2 compares the Grad-CAM distillation algorithm (decoupling common feature scores) with the distillation

TABLE 1 AP results for different models and after applying the knowledge distillation algorithm.

Method	mAP_50	Normal vibration damper	Vibration damper cross	Vibration damper corrosion	Normal single insulator	Insulator drop
Faster-Res101 (Teacher network)	0.544	0.648	0.312	0.642	0.589	0.530
Faster-Res50 (Student network)	0.419	0.556	0.286	0.370	0.423	0.458
Faster-Res101-KD-Res50	0.503	0.572	0.368	0.596	0.469	0.510
RetinaNet-Res101 (Teacher network)	0.559	0.699	0.069	0.576	0.751	0.701
RetinaNet-Res50 (Student network)	0.485	0.660	0.111	0.549	0.510	0.593
RetinaNet-Res101-KD-Res50	0.511	0.683	0.092	0.567	0.576	0.635
Cascade-101 (Teacher network)	0.617	0.672	0.464	0.695	0.631	0.623
Cascade-50 (Student network)	0.503	0.629	0.394	0.486	0.530	0.474
Cascade-101-KD-50	0.555	0.694	0.510	0.568	0.510	0.495

TABLE 2 Comparison of two foreground knowledge distillation methods.

Detection models	Settings	mAP_50	Normal vibration damper	Vibration damper cross	Vibration damper corrosion	Normal single insulator	Insulator drop
Faster R-CNN	Teacher network	0.544	0.648	0.312	0.642	0.589	0.530
	Student network	0.419	0.556	0.286	0.37	0.423	0.458
	Decoupled scoring	0.499	0.654	0.303	0.463	0.614	0.462
	Proposed method	0.503	0.572	0.368	0.596	0.469	0.510
RetinaNet	Teacher network	0.559	0.699	0.069	0.576	0.751	0.701
	Student network	0.485	0.660	0.111	0.549	0.510	0.593
	Decoupled scoring	0.518	0.723	0.101	0.554	0.582	0.630
	Proposed method	0.516	0.683	0.092	0.567	0.576	0.663

algorithm proposed in this article, using Faster R-CNN and RetinaNet as representatives of two-stage and single-stage detection models, respectively. The Grad-CAM distillation algorithm improves the detection performance of most targets, but for fittings defects, its improvement is slightly inferior to the distillation algorithm proposed in this article. In the case of the Faster R-CNN model, the detection accuracy for normal vibration damper targets and normal single insulator targets is slightly lower in this article's method compared to the Grad-CAM method. However, for vibration damper cross defects, our method

outperforms the Grad-CAM method by 6.5 percentage points, and for vibration damper corrosion and insulator String defects, it outperforms the Grad-CAM method by 13.3% and 4.8%, respectively. The analysis shows that the feature used for defect target detection is more concentrated in the critical pixel regions. Using ground-truth boxes as the range of foreground object regions introduces more noise for defects. This article's method selects the top-k highest-scored pixels to form a mask, which includes essential features for defect detection while avoiding introducing noise from other parts of the foreground object regions. [Figure 3](#)



TABLE 3 Ablation Experiments with Different k Values.

k	1	5	10	15	30	45	60	75
mAP_50	0.189	0.265	0.312	0.458	0.467	0.483	0.479	0.410

shows visualized detection images for different student networks, teacher networks, and networks after applying the distillation method.

3.3 Ablation experiments

To investigate the impact of different k values on improving the accuracy of the student network, ablation experiments were conducted based on the Faster-RCNN model, using ResNet-100 as the teacher network and ResNet-50 as the student network. The mAP_50 values of the student network were observed for different k values, and the results are shown in Table 3. From the results, it can be observed that the student network achieves the optimal mAP_50 value when k is set to 45. It is speculated that if k is too small, it may not capture crucial defect features, while

setting k to a large value introduces too much noise from foreground object regions, leading to negative effects.

4 Conclusion

Due to the limited dataset of hardware defects, hardware defects occur as damage to the hardware target. Compared to the hardware target, the effective feature range of hardware defects is smaller. In order to improve the detection accuracy of hardware defects in power transmission lines by the student network, this paper improves the foreground target region distillation. It guides the student network with more refined feature knowledge generated by the teacher network. Considering the influence of the two tasks, classification, and regression, in the foreground target region, pixels are scored, and the top k important pixels' generated masks containing feature knowledge are used to guide the student network. Experimental results show that the proposed method applied to three different single-stage and two-stage detection models, Faster-RCNN, RetinaNet, and Cascade R-CNN, has improved the detection accuracy of hardware and its defects. In Faster R-CNN, after adding the knowledge distillation algorithm in

this paper, mAP₅₀ has improved by 8.44% for the student network. RetinaNet improved by 2.6%, and Cascade R-CNN improved by 5.28%. This lays a solid foundation for lightweighting the hardware and its defects detection models in power transmission lines.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

ZZ: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing–review and editing. XL: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. YX: Writing–original draft, Writing–review and editing. SM: Formal Analysis, Supervision, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the National Natural Science Foundation of China

References

- Cao, Y., Xu, J., Lin, S., Wei, F., and Hu, H. (October 2019). “Gcnet: non-local networks meet squeeze-excitation networks and beyond,” in Proceedings of the IEEE/CVF international conference on computer vision workshops. Montreal, BC, Canada 0-0
- Du, Q., Dong, W., Su, W., and Wang, Q. (September 2022). “UAV inspection technology and application of Transmission Line,” in Proceedings of the 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), 594–597. Dalian, China. doi:10.1109/iciscae55891.2022.9927674.8
- Du, Z., Zhang, R., Chang, M., Zhang, X., Liu, S., Chen, T., et al. (2021). Distilling object detectors with feature richness. *Adv. Neural Inf. Process. Syst.* 34, 5213–5224. doi:10.48550/arXiv.2111.00674
- Gu, M., Ming, R., Qiu, C., and Wang, X. (2023). A multi-teacher knowledge distillation model compression algorithm for deep neural network. *Appl. Electron. Tech.* 49 (08), 7–12. doi:10.16157/j.issn.0258-7998.233812
- Guo, J., Han, K., Wang, Y., Wu, H., Chen, X., Xu, C., et al. (June 2021). “Distilling object detectors via decoupled features,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA 2154–2164.
- Hu, H., Gu, J., Zhang, Z., Dai, J., and Wei, Y. (June 2018). “Relation networks for object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA 3588–3597.
- Li, B., Qu, L., Zhu, X., Guo, Z., Tian, Y., Yu, F., et al. (2023). TMT proteomics analysis reveals the mechanism of bleomycin-induced pulmonary fibrosis and effects of Ginseng honeysuckle superfine powdered tea. *Trans. China Electrotech. Soc.* 38 (01), 60–70. doi:10.1186/s13020-023-00769-x
- Li, J., Wang, Q., Hong, S., Fan, L., Chen, X., and Ai, C. (June 2023). “Improved YOLOV7 algorithm for transmission line metal fittings equipment detection: algorithm improvement for integrating attention mechanism,” in Proceedings of the 2023 4th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China 496–500.
- Li, X., Li, Z., Wang, H., and Li, W. (2021). Unmanned aerial vehicle for transmission line inspection: status, standardization, and perspectives. *Front. Energy Res.* 9, 713634. doi:10.3389/fenrg.2021.713634
- Nguyen, V., Jenssen, R., and Roverso, D. (2018). Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning. *Int. J. Electr. Power & Energy Syst.* 99, 107–120. doi:10.1016/j.ijepes.2017.12.016
- (61871182 and U21A20486), the Natural Science Foundation of Hebei Province (F2020502009, F2021502008, and F20211502013), and the Fundamental Research Funds for the Central Universities (2023JC006).

Acknowledgments

Heartfelt thanks to everyone who contributed to this paper.

Conflict of interest

Author XL was employed by China Nuclear Power Engineering Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Signal and Information Processing, Networking and Computers (ICSINC), Singapore. 271–279.

Zhai, Y., Guo, C., Wang, Q., Zhao, K., Bai, Y., and Zhang, Y. (2023). Multi-fittings detection method for transmission lines based on implicit spatial knowledge fusion. *J. Graph.* 9, 1–10.

Zhang, K., Yue, Q., Huang, W., Huang, C., Ren, Y., and Hou, J. (2022). Clinical study of factors associated with pregnancy outcomes in pregnant women with systemic lupus erythematosus in the southern China. *Adv. Technol. Electr. Eng. Energy* 41 (11), 59–66. doi:10.46497/ArchRheumatol.2022.8870

Zhao, B., Wang, Q., Wu, Y., Cao, Q., and Ran, Q. (2022). Target detection model distillation using feature transition and label registration for remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 15, 5416–5426. doi:10.1109/jstars.2022.3188252

Zhao, Z., and Cui, Y. (2018). Research progress of visual detection methods for transmission line key components based on deep learning. *Electr. Power Sci. Eng.* 34 (03), 1–6. doi:10.3969/j.issn.1672-0792.2018.03.001

Zhao, Z., Jiang, Z., Li, Y., Qi, Y., Zhai, Y., Zhao, W., et al. (2021). Overview of visual defect detection of transmission line components. *J. Image Graph.* 26 (11), 2545–2560. doi:10.11834/jig.200689

Zhao, Z., Jin, C., Qi, Y., Zhang, K., and Kong, Y. (2021). Image classification of transmission line bolt defects based on dynamic supervision knowledge distillation. *High. Volt. Eng.* 47 (02), 406–414. doi:10.13336/j.1003-6520.hve.20200834

Zhao, Z., Xiong, J., Xu, H., and Zhang, L. (2023). Integrating structural reasoning for deep model transmission line fittings and their defects detection. *High. Volt. Eng.* 49 (08), 3346–3353. doi:10.13336/j.1003-6520.hve.20230241



OPEN ACCESS

EDITED BY

Hengrui Ma,
Qinghai University, China

REVIEWED BY

Linfei Yin,
Guangxi University, China
Anton Louise Pernes De Ocampo,
Batangas State University, Philippines

*CORRESPONDENCE

Xu Lu,
✉ bruda@126.com

RECEIVED 26 August 2023

ACCEPTED 23 October 2023

PUBLISHED 03 November 2023

CITATION

Liu J, Hu M, Dong J and Lu X (2023), The application of a lightweight model FA-YOLOv5 with fused attention mechanism in insulator defect detection. *Front. Energy Res.* 11:1283394. doi: 10.3389/fenrg.2023.1283394

COPYRIGHT

© 2023 Liu, Hu, Dong and Lu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The application of a lightweight model FA-YOLOv5 with fused attention mechanism in insulator defect detection

Jun Liu^{1,2}, Mingming Hu¹, Junyuan Dong¹ and Xu Lu^{1,3*}

¹Institute of Automation, Guangdong Polytechnic Normal University, Guangzhou, China, ²Guangzhou Key Laboratory of Intelligent Building Equipment Information Integration and Control, Guangzhou, China, ³Pazhou Lab, Guangzhou, China

Insulators are important components of transmission lines, serving as support for conductors and preventing current backflow. However, insulators exposed to natural environments for a long time are prone to failure and can cause huge economic losses. This article proposes a fast and accurate lightweight Fast and Accurate YOLOv5s (FA-YOLO) model based on YOLOv5s model. Firstly, attention mechanisms are integrated into the network module, improving the model's ability to extract and fuse target features. Secondly, the backbone part of the network is lightweighted to reduce the number of parameters and computations at the cost of slightly reducing the accuracy of detecting a few objects. Finally, the loss function of the model is improved to accelerate the convergence of the network and improve detection accuracy. At the same time, a visual insulator detection interface is designed using PyQt5. The experimental results show that the algorithm in this paper reduces the number of parameters by 28.6%, the computational effort by 35.7%, and the mAP value by 1.7% compared with the original algorithm, and is able to identify defective insulators quickly and accurately in complex backgrounds.

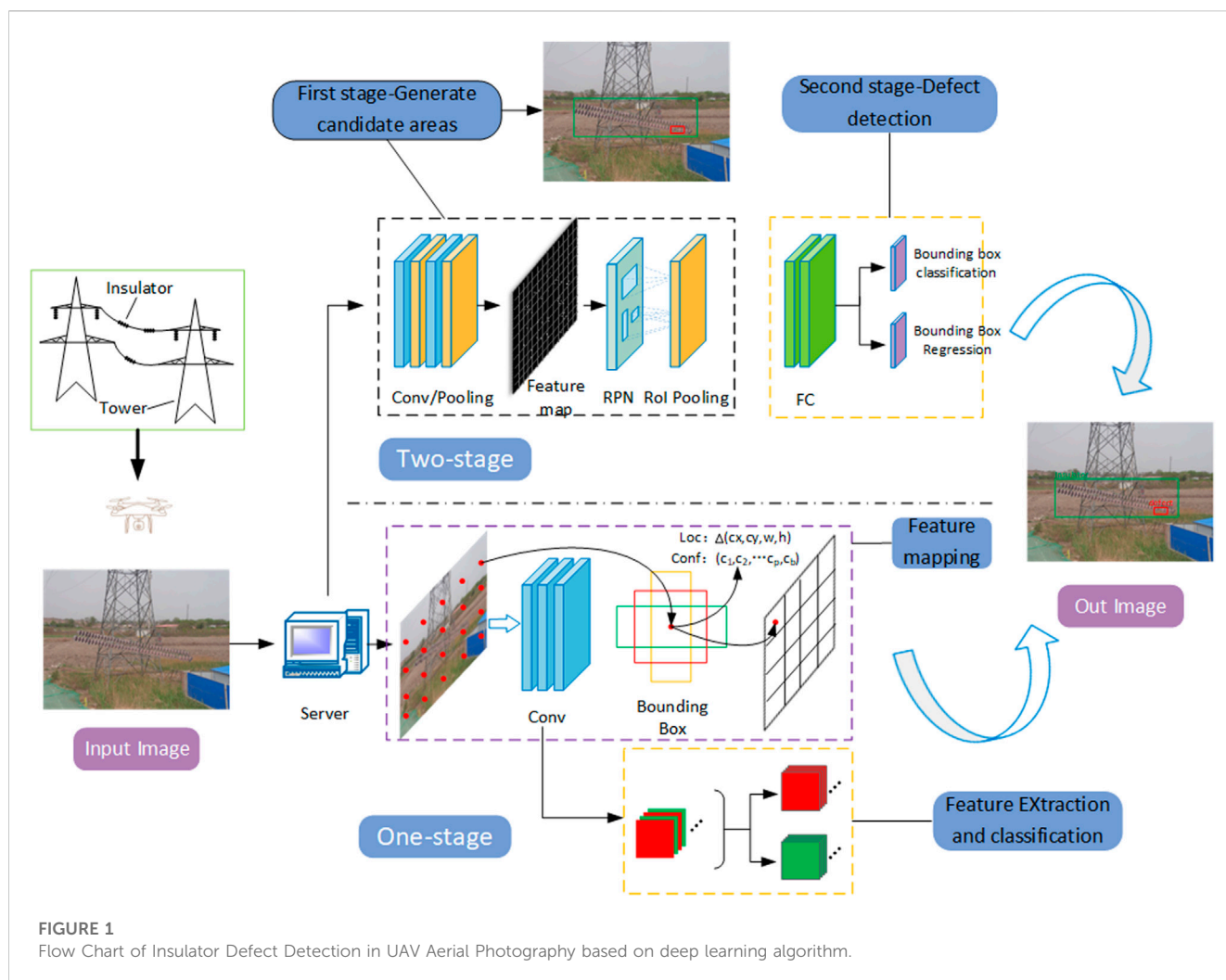
KEYWORDS

insulators, defect detection, attention mechanism, lightweighting, WIoU_Loss

1 Introduction

With the increasing demand for electricity, transmission lines have spread all over the country. Insulators, as an important part of them, have good mechanical support and electrical insulation properties, and play an important role in supporting the conductor and preventing the current from returning to the ground during the whole transmission process. However, its long-term exposure to strong electric field environment and susceptibility to adverse weather conditions such as rain, snow, and extreme temperatures, resulting in defects such as spontaneous explosion and fracture, creates a huge potential risk to the safe and stable operation of transmission lines, and according to statistics, the highest number of failures in power systems is caused by insulator defects (Chen, 2020; El-Hag, 2021). Therefore, fast and accurate detection of defective insulators and timely replacement are particularly important for the safe operation of the entire transmission system.

The defect detection of insulators is mainly divided into insulator localization as well as defect detection. The defect detection of insulators can be divided into manual observation, traditional image processing based and deep learning based methods. Among them, the manual observation method is time-consuming and labor-intensive, and has certain safety



risks (Yu et al., 2019). And the traditional image processing-based methods need to set the target features artificially, and different features need to be set for different targets, and the recognition accuracy is low, which can easily cause false detection or missed detection. In (Lu et al., 2017), an active contour model is proposed for insulator segmentation based on the shape and texture features of insulators, and the method proves to be effective in identifying defective insulators even in a cluttered background. Zhang et al. (2018) proposed a computer vision-based insulator feature extraction method, which extracts texture features through a grayscale co-occurrence matrix and then detects insulator features using local features. Although the traditional image processing-based method has been able to detect defective insulators well, the detection process is complicated and easily disturbed by the background environment, resulting in missed and false detection. To overcome the interference of complex background, Zheng H. et al. (2020) proposed an improved infrared insulator image detection model based on the complex substation environment, which improves the extraction capability for insulator infrared image features by generating new feature pyramids with feature enhancement modules. However, the infrared imaging-based method is susceptible to the influence of temperature leading to poor detection results.

The above methods based on traditional image processing can only accurately identify defective insulators in a specific environment because they cannot automatically extract insulator features, but insulators are usually in complex background environments such as rivers, farmlands, construction sites and forests, so a method that can automatically extract insulator feature information from images is urgently needed.

With the rise of deep learning technology, target detection algorithms have achieved great success in the field of insulator defect detection by virtue of their fast and accurate recognition capability. Compared with traditional image processing methods, deep learning-based target detection algorithms can automatically extract deep feature information in images, reduce recognition time and improve detection accuracy (Yang et al., 2021). The flow of insulator defect detection based on deep learning algorithm is shown in Figure 1.

At present, the mainstream object detection algorithms are mainly divided into Two stage and One stage. Two stage first generates a prior box based on the target object, and then recognizes and judges the objects within the prior box. This method has high detection accuracy and can accurately identify the target, but the detection time is long. The mainstream algorithms include Faster R-CNN (Ren et al., 2015) and Mask R-CNN (He et al.,

TABLE 1 Relevant datasets used to cite the Reference.

Author	Number of data	Method	Data enhancement	Fault location
Shuang et al. (2023)	806	Detail R-CNN	×	√
Zhao et al. (2021)	10,468	Faster R-CNN	×	√
Yi et al. (2023)	1700	YOLOv5s	√	√
Zhang et al. (2023)	848	YOLOv4	√	√
Miao et al. (2019)	6700	SSD	×	√
Chen Y et al. (2023)	2448	YOLOv8n	√	√

2017). For example, Shuang et al. (2023) introduced a feature enhancement and assisted classification module based on Faster R-CNN to improve the accuracy of model detection. The data enhancement method of YOLOv5-X was also ported to expand the dataset. Zhao et al. (2021) firstly used feature pyramid network to improve the Faster R-CNN model, and then segmented the image by hue, saturation and value color space (HSV) adaptive thresholding algorithm, and finally localized and detected the defective insulators. Tan et al. (2022) used Mask R-CNN model to segment out insulators and detected defects such as breakage, dirt, foreign matter and flashover by multi-feature fusion and cluster analysis model.

One stage directly locates and recognizes targets, which has a fast detection speed and can achieve real-time detection. However, the detection effect is not satisfactory. The mainstream algorithms include YOLO series algorithms and SSD (Wei et al., 2016) algorithm, and timely detection and replacement of defective insulators is important for the safe and stable operation of transmission lines. In order to detect the working status of insulators in real time, Yi et al. (2023) improved the Neck part of the YOLOv5s model and proposed a new attention module MainECA to enhance target perception, and the proposed YOLO-Small model reduced the number of parameters while improving the detection accuracy. Zhang et al. (2023) used GhostNet as the Backbone network of the YOLOv4 model, and at the same time optimized the model using K-means algorithm and Focal loss function. Chen Y et al. (2023) added the GSConv module to the latest YOLOv8n algorithm to reduce the complexity of the network, and also adopted a lightweight Content-Aware Feature Reconstruction (CARAFE) structure to enhance the feature fusion capability of the model. Miao et al. (2019) used a combination of SSD model and two-stage fine-tuning strategy to complete the detection of defective insulators, which can automatically extract multi-level features of images and can identify porcelain insulators and composite insulators quickly and accurately in complex backgrounds.

Based on the fact that deep learning methods need to use a large number of datasets to achieve better results, and then there are not many open-source insulator datasets due to confidentiality factors, most of the methods mentioned above use data augmentation strategies to expand their datasets, as shown in Table 1.

In Table 1, Shuang et al. used 806 images captured from Guangxi Power Grid in China as the dataset, and did not expand the dataset using image processing methods, but directly used the data enhancement methods in YOLOv5x to enhance the training data. Zhao et al. used 4 datasets with a total of 10,468 images, and Yi et al.

used data enhancement methods such as rotating, panning, scaling, cropping, etc. to expand 1700 original images to 5180 images. Zhang et al. also used the above methods to expand 848 images in the original open-source Chinese Power Line Insulator Dataset (CPLID) into 880 images. Zhang et al. also used the above method to expand 848 images in the original open-source Chinese Power Line Insulator Dataset (CPLID) to 5832 images, and Miao et al. used a drone to take 6700 original images on the transmission line as a dataset for their experiments. Finally, Chen et al. expanded the open-source datasets CPLID and Insulator Defect Image Dataset (IDID) to 5676 images using common data expansion methods.

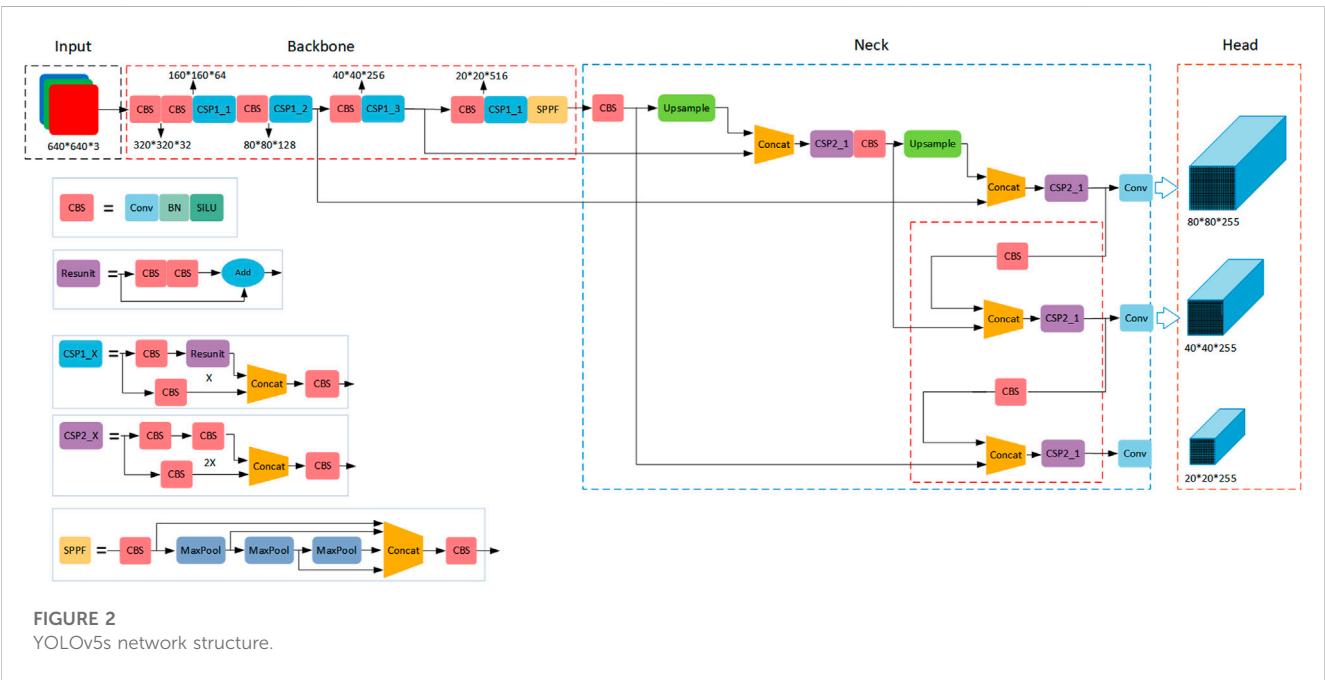
The main research objective of this article is to propose an improved YOLOv5 algorithm, FA-YOLOv5s, to address the issues of high computational complexity, slow detection speed, complex background, mutual occlusion, and small targets in the existing insulator defect detection algorithms. The proposed method mainly improves the network structure of YOLOv5 model and loss function, so that the new algorithm can quickly and accurately identify insulators in complex environments and detect whether they have faults. The main contributions of this article are as follows: 1) Integrating the Convolution Block Attention Module (CBAM) (Woo et al., 2018) with the network's C3 module enhances the network's ability to fuse insulation feature information, improving detection accuracy. 2) By using Partial Convolution (PConv) to lightweight the main network part of the model, the computation cost is reduced at the cost of reduced accuracy. 3) The loss function of the network was improved by using Wise_IoU Loss as the loss function, which improved the convergence speed of the model.

2 Related work

As one of the current popular target detection methods, the YOLOv5 algorithm is a product of continuous innovation and improvement based on the YOLOv3 (Redmon and Farhadi, 2018) and YOLOv4 (Bochkovskiy et al., 2020) algorithms. It combines the advantages of both algorithms, has fewer parameter quantities, and a simpler structure. While accelerating the detection speed, it also increases the detection accuracy, and achieves better detection results on PASCAL VOC (Everingham et al., 2015) and COCO (Lin et al., 2014) datasets. According to its network depth and width, YOLOv5 is successively YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, with the fastest detection speed and lowest accuracy. The comparison on the COCO datasets is shown in Table 2.

TABLE 2 Comparison of YOLOv5 parameters in COCO dataset.

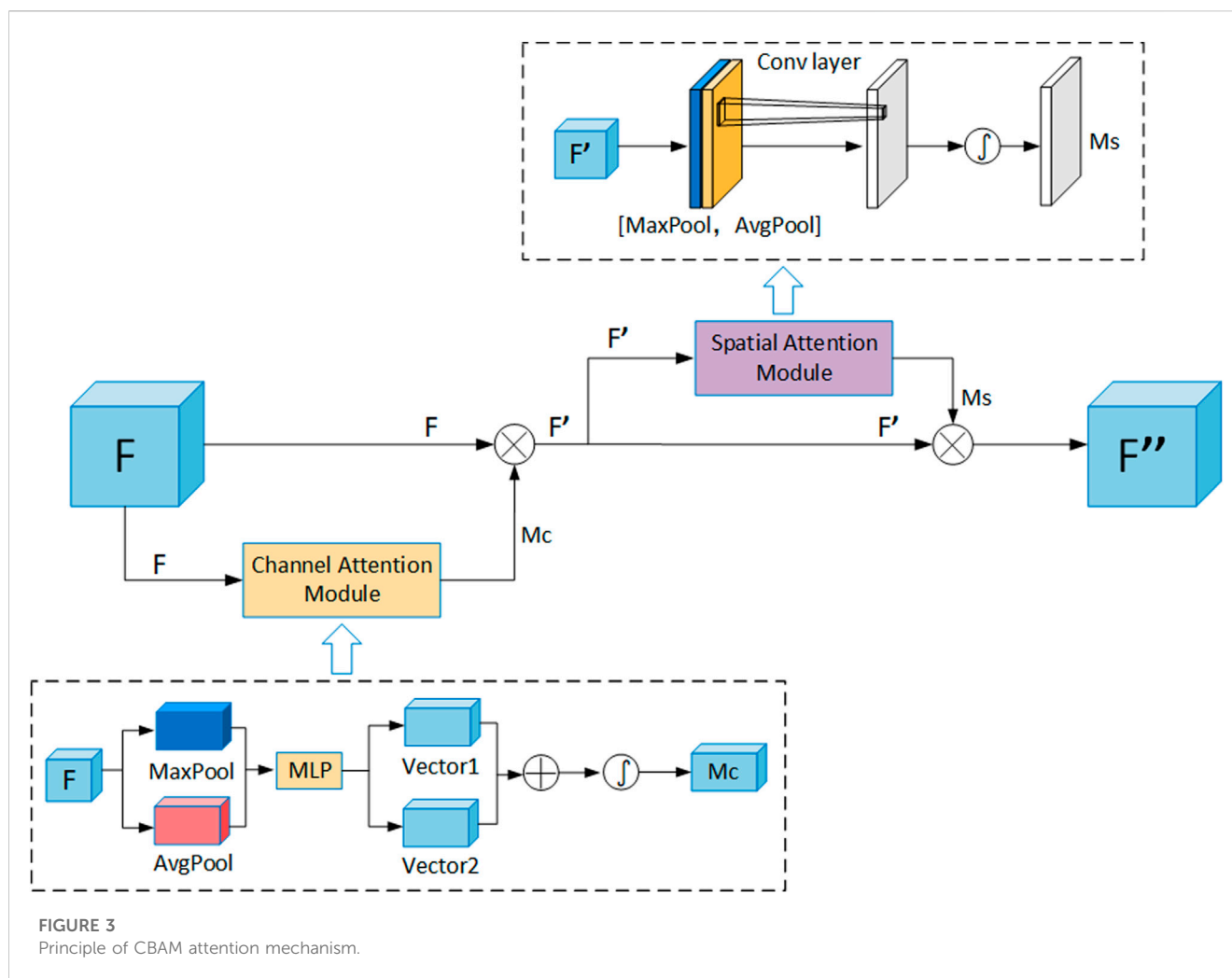
Network model	Image size	mAP 0.5	mAP 0.5:0.95	Speed (v100/ms)	Parameter(M)
YOLOv5s	640	55.4	36.7	2.0	7.3
YOLOv5m	640	63.1	44.5	2.7	21.4
YOLOv5l	640	66.9	48.2	3.8	47.0
YOLOv5x	640	68.8	50.4	6.1	87.7



The network model is shown in Figure 2. YOLOv5s network structure is mainly divided into four parts: Input, Backbone, Neck and Head. Mosaic (Lewy and Mańdziuk, 2023) data enhancement is performed on the input side to speed up the image processing and reduce the memory size of the model, which makes the model obtain better detection results. backbone is mainly composed of Conv-Batch Normalization-SiLU(CBS), Stack 3 convolutional layers on top of multiple bottleneck layers (C3) and Spatial Pyramid Pooling Fusion (SPPF) modules. The CBS module consists of a normal convolutional layer (Convolutional), a batch normalization (Batch Normalization) and an activation function (SiLU), while C3 consists of three standard convolutional layers and several Bottleneck modules, which are structured The SPPF is divided into two branches, one passing through multiple Bottleneck stacks and 3 standard convolutional layers, and the other passing through a basic convolutional module, and finally they are concatenated. SPPF improves the perceptual field of the network through feature extraction with maximum pooling of different pooling kernel sizes. The main role of the Neck part is to deep fuse the features extracted from the Backbone The Head part outputs the input size of 640 × 640 images as 20 × 20, 40 × 40 and 80 × 80 size feature maps, which are used to predict large, medium and small targets in three different sizes.

3 Fast and accurate FA-YOLOv5 algorithm

Current improvements to the YOLOv5 algorithm focus on improving the accuracy and convergence speed, while ignoring the complexity of the network model and the increase in the number of parameters, e.g., Han et al. (2022) added the ECA-Net attention mechanism to the backbone feature extraction network of YOLOv5, and also used a bidirectional feature fusion network in the feature fusion layer to enhance the detection of small targets. Gao et al. (2021) proposed a convolutional attention module with batch normalization (BN-CBAM) and a multi-level feature fusion module to enhance the detection of small targets. Although these methods are effective in improving the detection accuracy for small targets, they also make the network structure more complex and reduce the detection speed. In this paper, the convolutional attention mechanism CBAM module is fused with the C3 module of Neck part to improve the accuracy of detection. At the same time, the network is lightweighted to address the problems of complex network structure, number of parameters, and large computation. Finally, the latest WIoU loss is used as the loss function of the model, which speeds up the convergence, makes full use of the dynamic non-monotonic FM potential, and solves the problem of unbalanced sample quality.



3.1 Incorporating attention mechanism in C3 module

Since insulators are mostly in complex backgrounds and the defective part of insulators is a relatively small part of the whole image, it is difficult for the algorithm to extract feature information of insulators and their defects effectively. In order to enhance the extraction of target feature information, researchers proposed the attention mechanism (Vaswani et al., 2017), whose main role is to enhance the extraction of various appearance features of the target and make the algorithm biased to extract the features, the core of which is to make the network focus on the region of the target in the image rather than the whole image. By making the algorithm focus on the feature information of the target and ignore other unimportant information to improve the detection performance of the algorithm, the attention mechanism has been widely used in computer vision tasks such as target detection and image segmentation in recent years, and occupies an important position in the field of deep learning.

Attention mechanisms are usually divided into channel attention mechanisms and spatial attention mechanisms, which focus on the channel dimension and spatial dimension, respectively. Channel attention is used to deal with the

assignment relationship of feature map channels, while spatial attention allows neural networks to focus more on target regions in the image and ignore irrelevant regions, and simultaneous attention allocation to both dimensions enhances the effect of attention mechanisms on model performance.

The workflow of CBAM is shown in Figure 3, where the feature map is first passed through the channel attention module, then the feature map is multiplied with the channel weights and input to the spatial attention module, and finally the normalized spatial weights are multiplied with the feature map input to the spatial attention module to obtain the final weighted feature map. The final weighted feature map is obtained. This module not only saves parameters and computational effort, but also can be easily added to other network structures. For example, Wang et al. (2022) directly added the CBAM attention module to the YOLOv5s network structure to improve the insulator feature extraction capability and achieve insulator detection in complex backgrounds, but the method is not effective for insulator defect detection of small targets.

The overall formula of Figure 3 is shown in Eqs 1, 2.

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (1)$$

$$M_s(F) = \sigma(f^{\text{conv}}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (2)$$

TABLE 3 Comparison of different attention mechanisms and addition methods.

Methods	Precision (%)	Recall (%)	mAP0.5 (%)	Parameters(M)	GFLOPs
YOLOv5	95.5	94.5	95.4	7.0	16.0
NAM	96.1	94.0	95.2	7.0	16.0
CBAM	96.5	94.7	95.9	7.0	16.0
ECA	95.4	95.3	95.5	7.0	16.0
CoordAtt	96.8	95.7	96.3	7.0	16.0
GAM	96.7	95.1	96.4	7.5	16.3
C3CBAM	96.7	96.1	96.5	6.9	15.7

The bold portion of the table indicates the value with the best performance in the metric.

$M_c(F)$ and $M_s(F)$ are the channel and spatial attention weights of the feature layer F , respectively; AvgPool and max Pool are the average pooling and maximum pooling operations; MLP stands for multilayer perceptron; σ is the sigmoid activation function. as the input to the next layer, as shown in Eq. 3.

$$\begin{cases} F' = M_c(F) \otimes F \\ F'' = M_s(F') \otimes F' \end{cases} \quad (3)$$

where \otimes denotes element-wise multiplication, F is the intermediate quantity of the feature layer passing through the channel attention module. and F'' denotes the output passing through the spatial attention module.

The current common method for improving attention mechanisms is to directly add them to the network structure, which does not fully leverage the effectiveness of attention mechanisms. Although this approach does improve the detection accuracy to some extent, it also increases the number of layers and parameters in the network. To further reduce the number of model parameters and fully leverage the effectiveness of attention mechanisms, this article combines attention mechanisms with the C3 module to form a new module, namely, C3CBAM. At the same time, the newly generated C3CBAM module further enhances the model's capability to focus on target feature information. This module strengthens the model's ability to fuse and extract target feature information from both channel and spatial dimensions, allowing for accurate identification of target feature information even in complex background environments. As a result, efficient insulation defect detection can be achieved.

To verify the effectiveness of this method, we conducted experiments by adding different attention mechanisms after the same C3 layer and compared them with the method of incorporating CBAM into the network layer. The experimental results are shown in Table 3. We added different attention mechanisms to the network model for comparison experiments, which were conducted after adding different attention mechanisms to the same Conv layer while ensuring that the number of other parameters of the experiment was the same. As can be seen from Table 3, different methods have different effects on the performance of the original model, Normalization-based Attention Module (NAM) and Efficient Channel Attention (ECA) reduce the accuracy of the model detection. The other attention mechanisms all have some improvement effect on the detection performance of the model, among which the Global Attention Mechanism (GAM) attention

mechanism has the biggest improvement effect, but it increases the number of parameters and computation of the model, because the purpose of this study is for fast and accurate insulator defect detection algorithm, out of the comprehensive considerations, we choose to integrate the CBAM Attention Mechanism and C3 module fusion method to improve the original model. This method reduces the number of parameters and computation to some extent, and most importantly has the highest performance enhancement effect on the original model.

3.2 Lightweight network architecture

The FasterNet (Chen J et al., 2023) network recently released by CVPR far exceeds other existing networks in terms of lightweight as well as the balance of detection performance. The current mainstream lightweight networks such as MobileNet, ShuffleNet, and GhostNet utilize deep convolution (DWConv) or group convolution (GConv) to extract spatial feature information, which although greatly reduces the number of parameters and floating point operations (FLOPs), but the computation is not efficient, increases the number of layers of the network, runs slower, and greatly reduces the accuracy and effectiveness of detection, while adding some additional data operations. In order to maintain high accuracy while reducing FLOPs, Chen et al. proposed local convolution (PConv), which works as shown in Figure 4.

PConv has lower computational effort as well as higher computational efficiency, which can utilize the computational power of the device more efficiently and also improves the model's ability to extract spatial feature information. Based on this, Chen Y et al. (2023) proposed FasterNet, which can achieve better results in classification, detection and segmentation tasks at a faster rate, and its can replace the Backbone part of the YOLOv5 model.

This article improves the backbone network of YOLOv5 using PConv, FasterNet, MobileNet, ShuffleNet, and GhostNet respectively. Through experimental comparisons, it is shown that PConv effectively reduces the complexity and parameter count of the network while maintaining high accuracy. The experimental results are shown in Table 4. The backbone network using the FasterNet improvement algorithm has the highest detection accuracy but the number of parameters is still high, while replacing the entire backbone part of the network using ShuffleNetV2 greatly reduces the number of parameters and computation, but also increases the number of layers of the network, and the detection speed is also reduced.

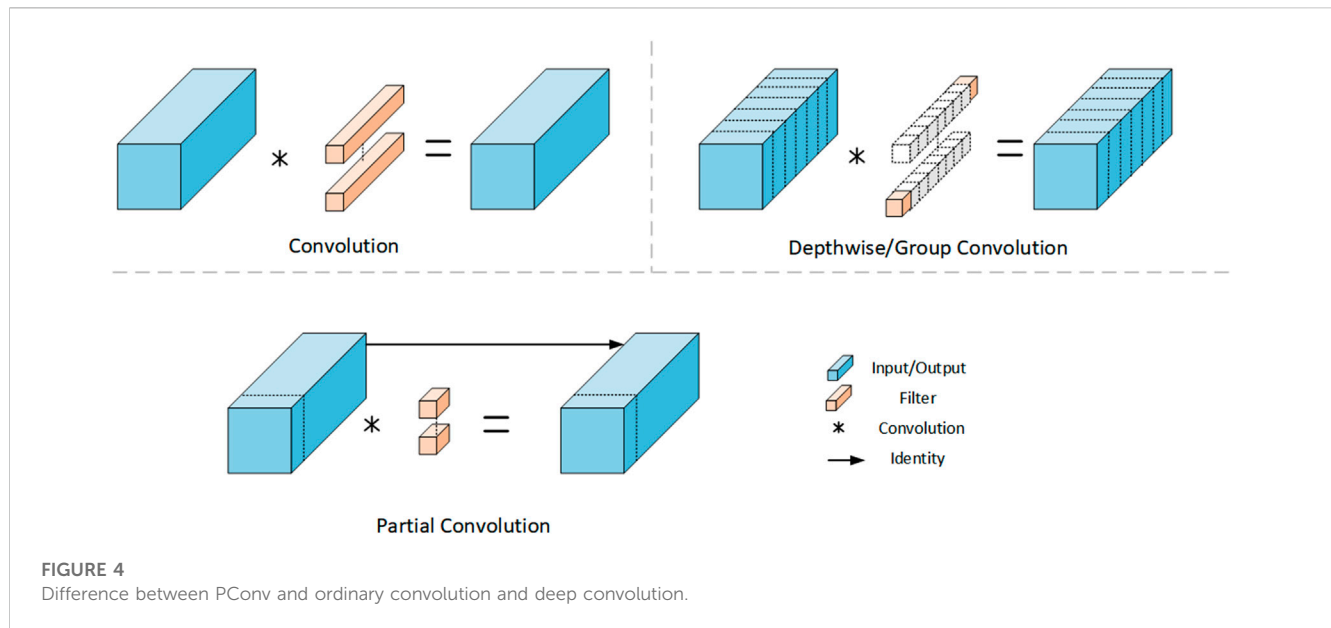


TABLE 4 Experimental results of different ways of light weight treatment.

Methods	Layers	Parameters(M)	GFLOPs	Precision (%)	mAP0.5 (%)	FPS
PConv	129	5.1	10.5	95.1	94.4	89
Faster-Net	228	6.4	14.0	95.4	95.2	76
MobileNetV3	320	1.4	2.3	94.8	93.2	80
ShuffleNetV2	193	0.8	1.9	94.5	92.9	82
GhostNet	500	5.3	8.4	96.8	94.1	73

The bold portion of the table indicates the value with the best performance in the metric.

MobileNetV3 and GhostNet both reduce the number of network parameters and computation at the cost of increasing the number of network layers. Based on this, this paper uses PConv to lighten the Backbone of the network, which can greatly reduce the number of parameters and computation of the model at the cost of a small reduction in accuracy.

3.3 Improvement of the loss function

The loss functions of YOLOv5 model are Classification loss, Localization loss and Confidence loss, and the sum of the three loss functions is the size of the total loss function. The calculation formula is as in Eq. 4.

$$\text{Loss} = \text{loss}_{\text{box}} + \text{loss}_{\text{obj}} + \text{loss}_{\text{cls}} \quad (4)$$

IoU_Loss (Yu et al., 2016) is the first proposed loss function for target detection, but it only considers the overlap area of the detection frame and the target frame, which has certain defects. the appearance of GIoU_Loss (Rezatofighi et al., 2019) loss function solves the shortcomings of IoU_Loss to a certain extent, but it also has the disadvantages of not accurate enough boundary regression and slow convergence speed. The subsequent DIOU_Loss (Zheng Z.

et al., 2020) loss function takes the overlap area and centroid distance into account and accelerates the convergence speed, but does not take the aspect ratio factor into account. To address these drawbacks, the CIOU_Loss (Zheng et al., 2021) loss function takes into account the overlap area, centroid distance and aspect ratio influence factor α and v , and its calculation process is shown in Eq. 5.

$$\text{CIOU} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{C^2} + \alpha v \quad (5)$$

where, b represents the center coordinates of the prediction frame, b^{gt} represents the parameter of the center of the real target bounding box. ρ^2 represents the Euclidean distance between the two centroids, c represents the minimum external rectangle diagonal length of the two rectangles, v is used to measure the consistency of the aspect ratio, and α is the weight function. The values of α and v are shown in Eq. 6.

$$\begin{cases} \alpha = \frac{v}{(1 - \text{IoU}) + v} \\ v = \frac{4}{\pi^2} \left(\arctan \frac{m^{\text{gt}}}{n^{\text{gt}}} - \arctan \frac{m}{n} \right)^2 \end{cases} \quad (6)$$

The original YOLOv5 algorithm uses CIOU_Loss as the loss function of the network, however, the v -value used to measure the

aspect ratio is too complex and slows down the convergence to some extent, so when one of the two variables increases (shrinks), the other one will shrink (increases). To solve this problem, (Zhang et al., 2022), proposed EIoU Loss by splitting the aspect ratio on the basis of CIoU, which accelerated the speed of convergence and improved the accuracy of regression. Focal-EIoU was also proposed to focus on high-quality anchor frames, which optimized the problem of sample quality imbalance in the regression task and made the regression process more focused on high-quality anchor frames, and the calculation process of EIoU_Loss is shown in Eq. 7. Thus (Yang et al., 2022) used EIoU as a loss function to improve the YOLOv3 algorithm, which improved the overlap between the predicted and actual frames of the target and accelerated the convergence speed.

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} \\ = 1 - IoU + \frac{\rho^2(b, b^{gt})}{C^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (7)$$

Although Focal-EIoU solves the problem of sample quality imbalance to some extent, the potential of non-monotonic FM is not fully utilized due to its static focusing mechanism (FM), so (Tong et al., 2023) proposed an IoU-based loss with dynamic non-monotonic FM, namely, Wise IoU (WIoU), which has a bounding box regression of attention-based loss WIoU v1, WIoU v2 with non-monotonic FM, and WIoU v3 with dynamic non-monotonic FM. In this paper, WIoU v3 is used as the loss function of the network, and its gradient gain allocation strategy with dynamic non-monotonic FM is utilized to trade-off the learning ability of high quality as well as low quality samples and improve the overall performance of the model. The calculation formula is shown in Eq. 8.

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \\ R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 - H_g^2)^*}\right) \quad (8)$$

where W_g , H_g denote the width and height of the minimum enclosing frame. To prevent R_{WIoU} from creating gradients that hinder convergence, W_g and H_g are separated from the computational graph (the superscript * indicates this operation). No new metric like aspect ratio is introduced because it effectively eliminates the factors that hinder convergence. To significantly amplify the localization loss (LIoU) of the normal quality anchor box, the range of R_{WIoU} is [1,e] while the range of LIoU is [0,1], which will significantly reduce the Rwiou of the high quality anchor box and focus on their centroid distance when the anchor box overlaps with the target box.

3.4 Network structure of this paper

In order to reduce the complexity of the model and make it more suitable for deployment on mobile devices such as UAVs, this paper uses PConv to lighten the backbone part of the network. At the same time, CBAM attention is fused with C3 module to give full play to CBAM's ability to extract target feature information in channel and space, which improves the accuracy of detection. Finally, WIoU_loss is used as the loss function of the network, and the improved part is shown in red, and the specific network structure is shown in Figure 5.

- (1) Backbone: Compared with the old version of Spatial Pyramid Pooling Fast (SPP) (He et al., 2015), the new version uses Fast-SPP (SPPF) to improve the processing speed of feature information. And replace all C3 modules in the backbone network with Pconv reduces the number of parameters as well as the computational effort of the network model, making the model able to run on low performance servers and more suitable for deployment on mobile devices.
- (2) Neck: This part mainly consists of Feature Pyramid Networks (Lin et al., 2017) and Perceptual Adversarial Network (Liu et al., 2018), which first fuses the input insulator feature maps from top to bottom to transfer the semantic information from the deep layer to the shallow layer to enhance the semantic representation at multiple scales, and then performs a bottom-up feature fusion to transfer the location information from the bottom layer to the deep layer to enhance the localization at multiple scales. The fusion of C3 module with CBAM attention mechanism in this part strengthens the ability of Neck part for fusion of target feature information, especially for small target insulators and self-detonation defective parts of insulators, and also reduces the complexity of the model to some extent.
- (3) Head: This part mainly detects 3 different scales, including some convolutional layers, pooling layers and fully connected layers, etc. Its role is to perform multi-scale target detection on the feature maps extracted from the backbone network. The model proposed in this article uses WIoU loss to improve detection accuracy and convergence speed in this section.

In order to verify the effectiveness of the method in this paper, the heat map visualization operation (Quan et al., 2022) was performed on the insulator feature extraction process in complex backgrounds, as shown in Figure 6, from which it can be seen that after the convolutional layer extracts the shallow information of insulators, the model can effectively segment the region where the target is located from the background environment; the sampling effect is obviously enhanced after the second stage C3CBAM feature extraction; after the third and fourth stage processing, the higher-level semantic information of the feature map has been more blurred, and the extracted insulator features have become abstracted. From the visualization results of the heat map, it is clear that the algorithm of this paper can more fully extract the color, texture, shape and edge information of insulator defects in the image, so as to quickly and accurately detect defective insulators.

Meanwhile, we designed a visual detection interface based on PyQt5 for the algorithm in this paper, as shown in Figure 7, which mainly has the following functions:

Model, select different models. Input, select the files to be detected, including the detection of pictures and videos in local files, and also has the function of real-time detection using the device's camera and supports RTSP video streaming. The ability to adjust the IoU, confidence level and frame rate delay in the detection process of the model, when reducing the IoU and confidence level, can make the model detect more targets, but the detection error is higher. When IoU and confidence are adjusted up, the accuracy of detection increases and the rate of missed detection increases. The delay can also be selected independently during the detection process. The interface also has the functions of start, pause and end, and the detection results are counted at the bottom left of the interface, and the results are automatically saved when the detection is completed.

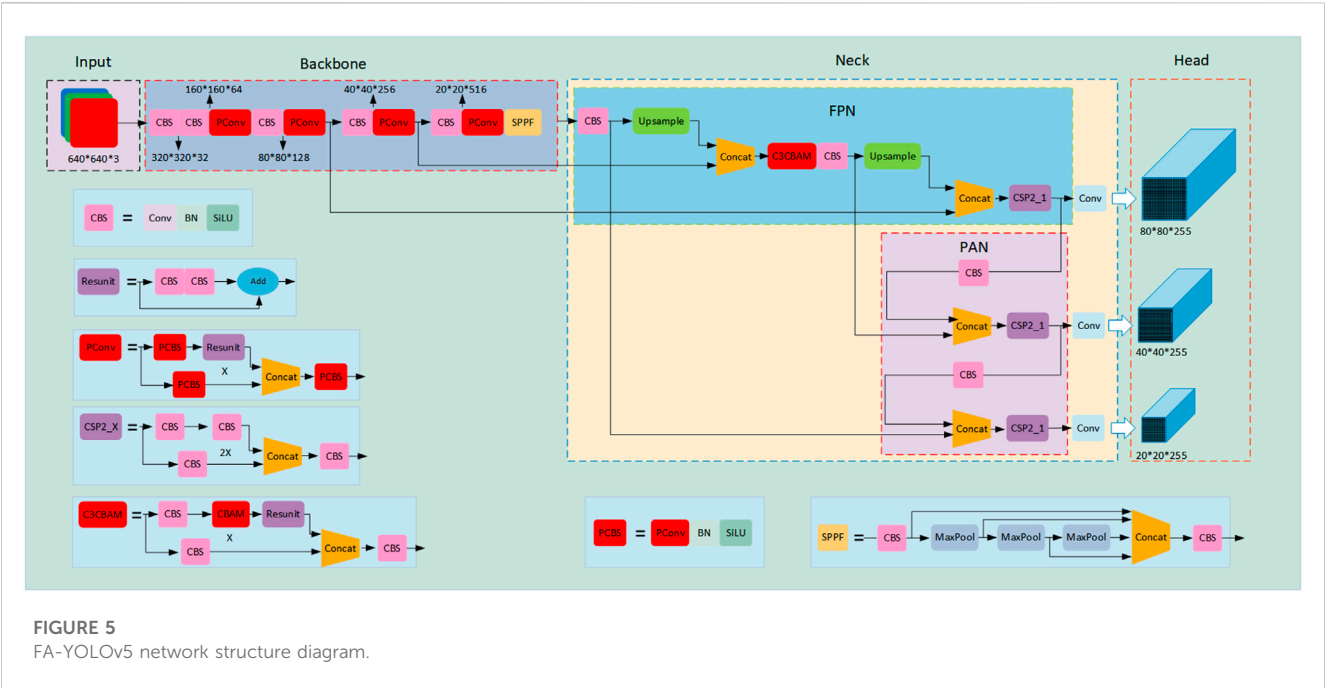


FIGURE 5
FA-YOLOv5 network structure diagram.

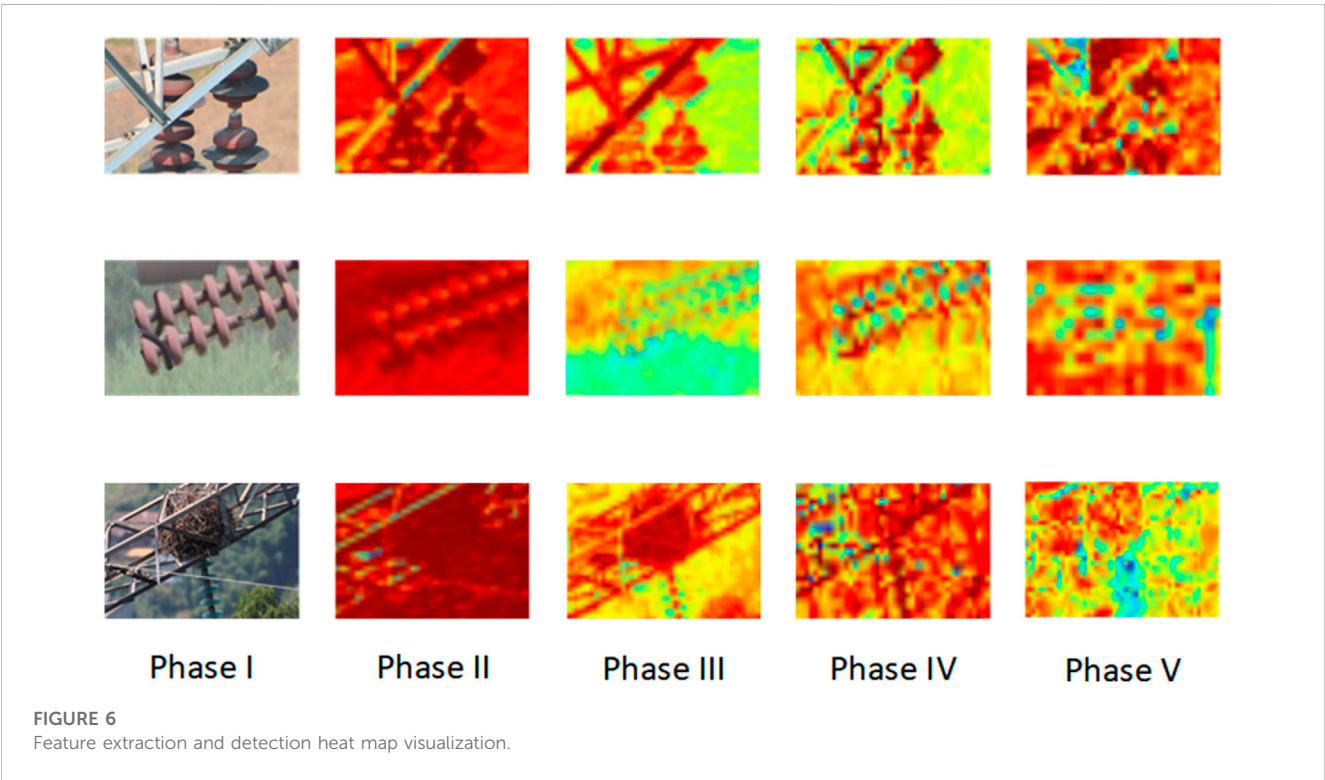


FIGURE 6
Feature extraction and detection heat map visualization.

4 Experimental results and analysis

4.1 Experimental environment and data pre-processing

The operating system for the experiments in this paper is Window 11, the CPU model is Intel(R) Core(TM) i7-11700 2.5GHz, 64GB RAM,

and the GPU model is GeForce RTX 3060 Laptop GPU with 12G video memory size of the workstation. The experimental environment is Python 3.8, GPU acceleration software CUDA 11.1 and CUDNN 8.1.0. The datasets used in this paper is mainly derived from three parts, with a total of 1006 insulator images. The first part is the Chinese power line insulator datasets (CPLID) (Raimundo, 2020), which includes 600 images of normal insulators and 248 images of self-exploding

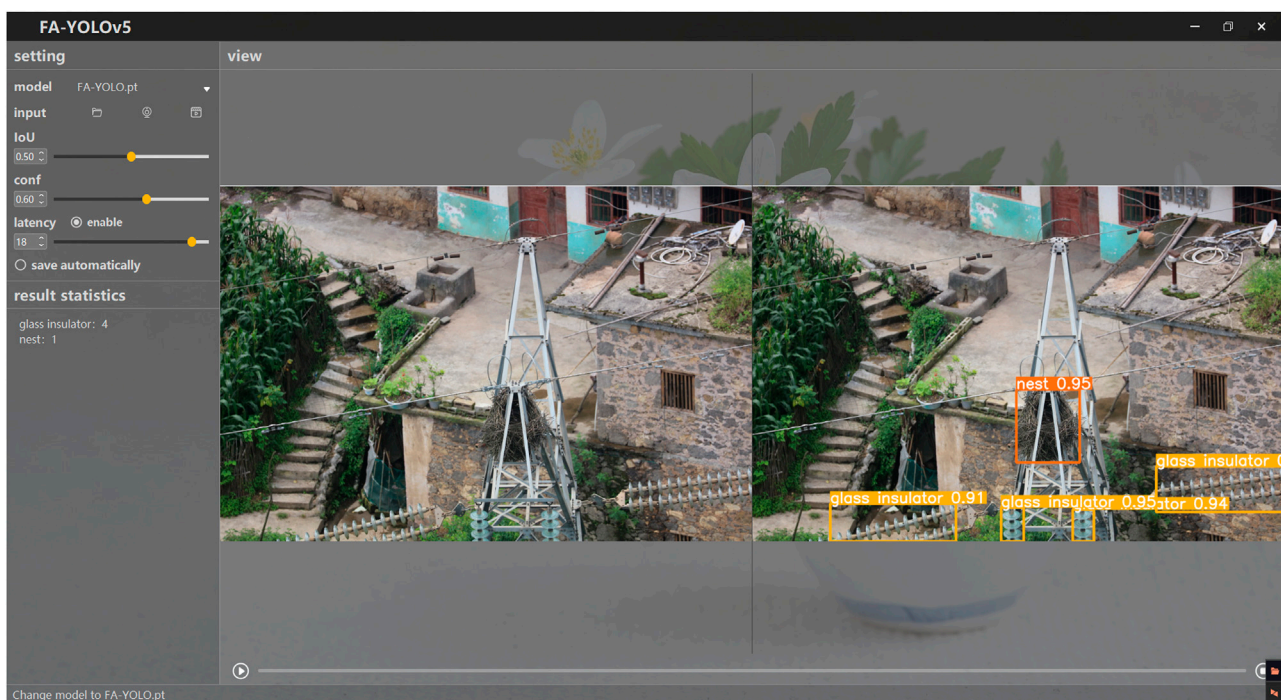


FIGURE 7
PyQt5-based visual inspection interface.

insulators. The second part is 40 self-exploding images of glass insulators disclosed by Baidu Flying Paddle; the third part is 118 images containing self-exploding glass insulators as well as bird's nests taken on site in a southern power grid. The data set was also labeled by LabelImg software, and the labeling labels were divided into: normal insulator (insulator), self-detonation defect (defect), bird's nest (nest) and glass insulator (glass insulator). Due to the lack of sufficient number of datasets, we expanded the number of datasets to 5174 by Gaussian blurring, cropping, brightness variation, and flipping of the existing datasets, and the results of partial data enhancement are shown in Figure 8. And the ratio of training set, validation set and test set is divided randomly in the form of 8:1:1. The input image size is 640×640 , the batch size is 16, the initial learning rate is 0.001, the network parameters are updated using SGD, the learning momentum is 0.937, the weight decay is 0.0005, warmup momentum is 0.8, the translate parameter is set to 0.1, and each training is 100 epochs.

4.2 Evaluation metrics

In order to accurately evaluate the performance of the algorithm, Precision (P), Recall (R), Average Precision AP and Mean Average Precision (mAP) are the most commonly used model evaluation metrics in the field of target detection, which are calculated as shown in Eqs 9–12, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$\text{AP} = \int_0^1 P(r)dr \quad (11)$$

$$\text{mAP} = \frac{\sum_{i=1}^N \text{AP}_i}{N} \quad (12)$$

Where TP denotes the number of positive samples predicted as positive by the model, FP denotes the number of negative samples predicted as positive by the model, i.e., false detection, and FN denotes the number of positive samples predicted as negative by the model, i.e., missed detection. N is the total number of detected categories, and in this paper N is set to 4, i.e., normal insulators, self-detonation defective insulators, bird's nests, and glass insulators. AP is the area enclosed by the PR curve, mAP is the detected average value of AP for each category. The larger the mAP, the better the performance of the algorithm.

4.3 Ablation experiment

In order to verify the effectiveness of the algorithm proposed in this paper, mAP, Precision, Recall, parameter quantity, and FPS were used as evaluation indicators to compare the performance of the model through ablation experiments. A total of 6 sets of models were used. Group A is the original datasets for YOLOv5s model training, Group B is the expanded datasets for YOLOv5s model training, Group C, D and E add PConv, C3CBAM and WIoU loss function respectively on the basis of Group B, and Group F (Ours) add PConv, C3CBAM and WIoU loss function on the basis of Group B, and carry out comparative

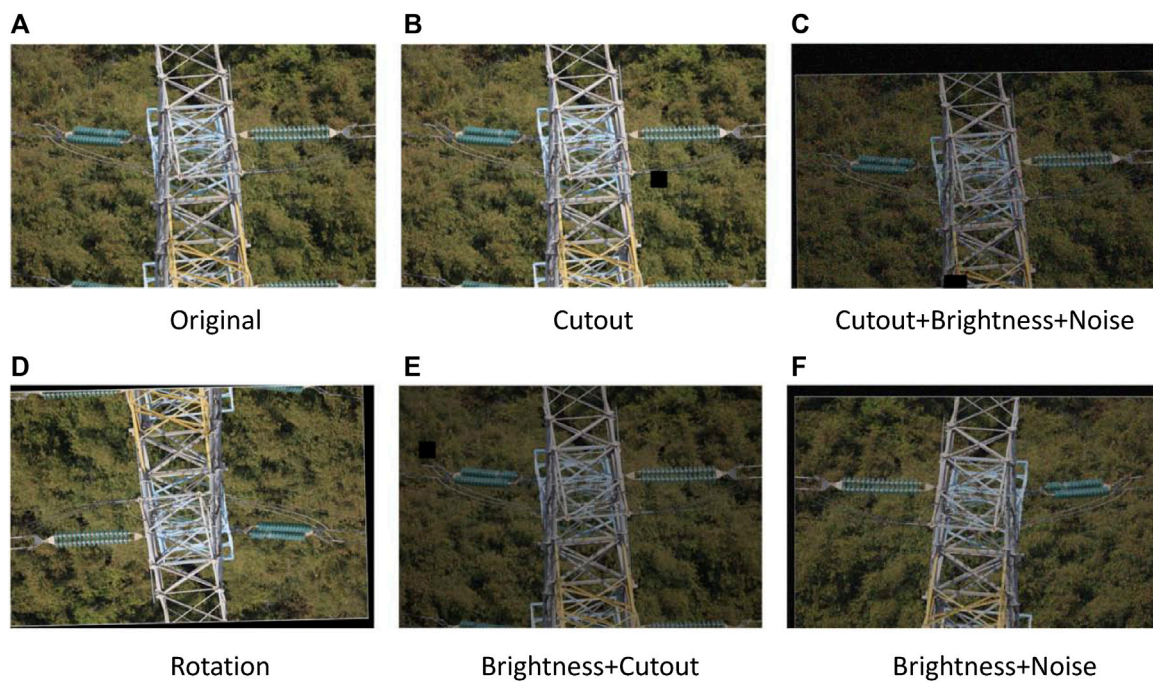
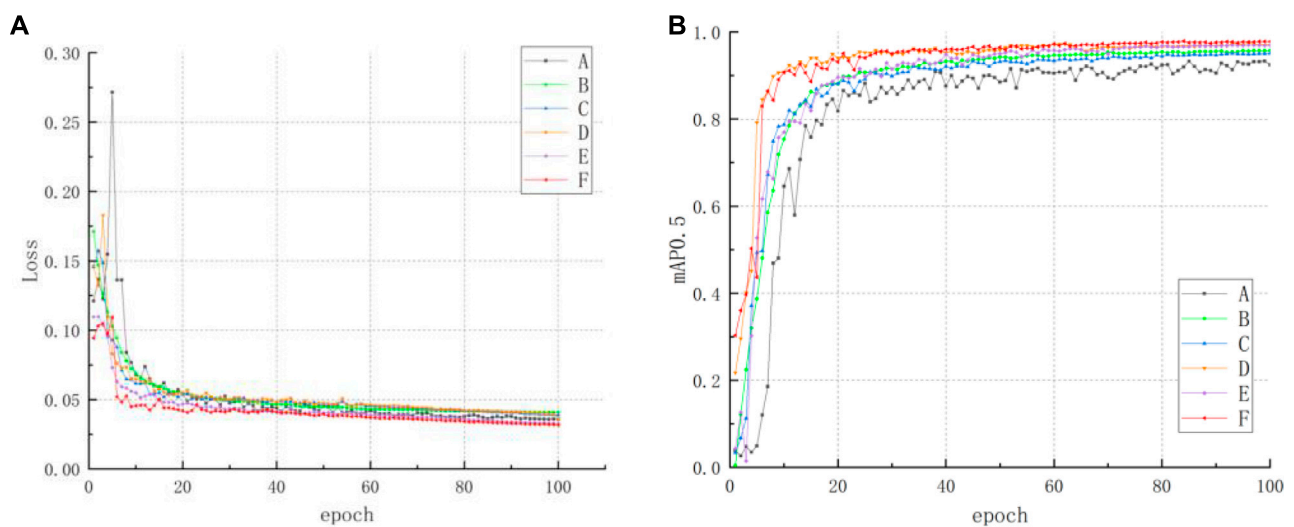


FIGURE 8
Partial data enhancement results.



mAP0.5 curve based on FA-YOLOv5 ablation experiment.
experiment.

FIGURE 9
Based on FA-YOLOv5 ablation experimental graph.

experiments with the same parameters. The experimental results are shown in Figure 9 and Table 5, respectively.

As can be seen from Table 5, before data enhancement, the YOLOv5 algorithm was not effective in detecting insulator defective parts due to the lack of sufficient defective samples, and after the data enhancement operation, it can be seen that the algorithm has

significantly improved the detection accuracy for all four categories, but there is still some room for improvement in the number of parameters, computation and overall performance of the model. When we use PConv to improve the backbone network part of the model, the number of parameters of the model is reduced by 27.2% and the computation is reduced by 34.4%, while the speed of

TABLE 5 Ablation experiments based on improved FA-YOLOv5.

/	Model	Precision (%)				Recall (%)	mAP0.5 (%)	Parameters(M)	GFLOPs	FPS	Weights (MB)
		Insulator	Defect	Nest	Glass insulator						
A	YOLOv5(Original datasets)	90.2	94.8	87.8	96.7	85.1	92.4	7.0	16.0	81	13.7
B	YOLOv5(Data Enhancements)	94.6	95.5	95.0	97.2	94.5	95.4	7.0	16.0	81	13.7
C	YOLOv5(Data Enhancements + PConv)	93.8	94.7	94.2	96.7	94.0	94.4	5.1	10.7	89	10.0
D	YOLOv5(Data Enhancements + C3CBAM)	95.9	96.3	96.8	96.4	96.1	96.5	6.9	15.7	80	13.6
E	YOLOv5(Data Enhancements + WIoU)	95.6	97.2	97.5	96.9	96.5	95.9	7.0	16.0	85	13.7
F	YOLOv5(Data Enhancements + PConv + C3CBAM + WIoU)	95.4	97.6	97.2	97.3	96.7	97.1	5.0	10.3	89	9.83

The bold portion of the table indicates the value with the best performance in the metric.

TABLE 6 Comparison of experimental results of different algorithmic models.

Methods	Precision (%)	Recall (%)	mAP0.5 (%)	Parameters(M)	GFLOPs
YOLOv5s	95.5	94.5	95.4	7.0	16.0
YOLOv7	96.9	96.0	96.1	37.2	105.2
YOLOv8s	97.3	96.8	96.6	11.1	28.7
SSD-VGG	85.6	67.8	84.3	26.3	62.7
Faster R-CNN	82.1	70.1	81.6	137.1	370.2
FA-YOLOv5	97.6	96.7	97.1	5.0	10.3

The bold portion of the table indicates the value with the best performance in the metric.

detection is improved to some extent, but the accuracy of detection is reduced. The fusion of the CBAM attention mechanism with the C3 module not only improves the detection accuracy but also reduces the complexity of the network. To further improve the performance of the model, we use Wise_Loss as the loss function of the model. Finally, a faster and more accurate model FA-YOLOv5 is proposed, which has 1.6% higher mAP value, 28.6% lower number of parameters, and 35.7% lower computational effort compared to the original model.

The comparison graph of experimental results is shown in Figure 9. Analysis of the mAP0.5 graph in Figure 9A shows that the convergence of the original algorithm is slow and the accuracy is low when no data augmentation is performed. After the data enhancement of the defective samples, the situation is significantly improved, and it can also be seen that the algorithm of this paper has stabilized at the 40th round and achieved a high detection accuracy. From the Loss plot in Figure 9B, it can be seen that the loss value of the algorithm for training the original datasets only starts to

stabilize in the 53rd epoch, while the loss value of the algorithm after doing data enhancement operation on the original datasets slowly stabilizes after 30 epochs of training, but the loss values of the algorithm are all improved, and it can be seen from the curve sets of Group E (WIoU) and Group F (Ours) that the improvement of the loss function in this paper has obvious effect on speeding up the convergence, while the loss value reaches the minimum.

4.4 Comparison experiments

To further verify the superiority and feasibility of the algorithm in this paper, we conducted comparison experiments on the unimproved YOLOv7, YOLOv8s, SSD and Faster R-CNN algorithms with optimal parameters, and the datasets used for the experiments were all self-built insulator defect datasets in this paper, and the precision, recall and average precision during the experiments of the mean value are shown in Table 6.

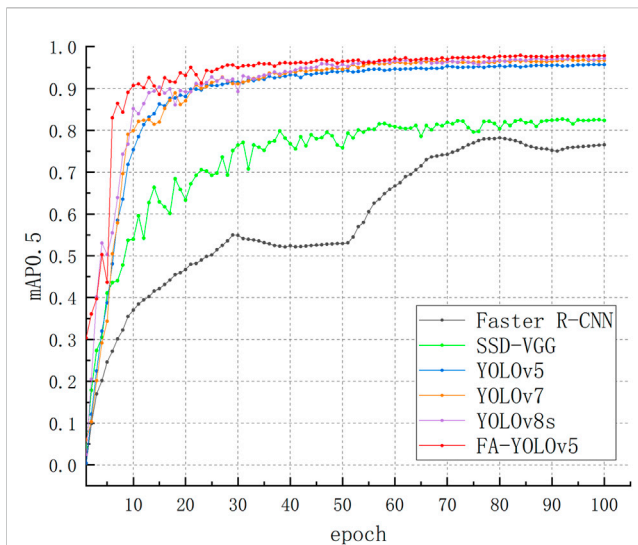


FIGURE 10
Comparison of experimental results of mAP0.5 curves for different models.

From Table 6, it can be seen that among the unimproved algorithm models, YOLOv8s and YOLOv7 models, as the latest target detection algorithms nowadays, have high detection

accuracy, but compared with YOLOv5s, their number of parameters and computation amount are larger. Faster R-CNN, as a typical Two-stage algorithm, has the highest number of parameters and the largest computation amount, and also the worst performance among all the compared algorithm models. SSD algorithm, as one of the typical One-stage algorithms, has only a little bit more parameter and computation amount than the YOLOv5 algorithm, but due to the fact that the last layer of the feature map of the network structure is too small, it is easy to lose the feature information of the target, which leads to the loss of feature information of the target. which leads to easy loss of the target's feature information, so the detection effect for this dataset is also poor. In order to balance the detection accuracy and model complexity, this paper proposes a lightweight model FA-YOLOv5 with better detection performance on the basis of YOLOv5 model, which has the highest detection accuracy and the least network parameters and computation among the listed models, and it is more suitable for deploying on mobile devices for transmission line inspection such as UAVs, which proves the feasibility of the method in this paper. Meanwhile, from the mAP0.5 curve graph in Figure 10, it can also be more intuitively seen that the algorithm proposed in this paper has a better convergence speed, and at the same time, it also has a better detection accuracy, and its detection performance is better than that of other comparative algorithms, which further proves the effectiveness of the algorithm in this paper.

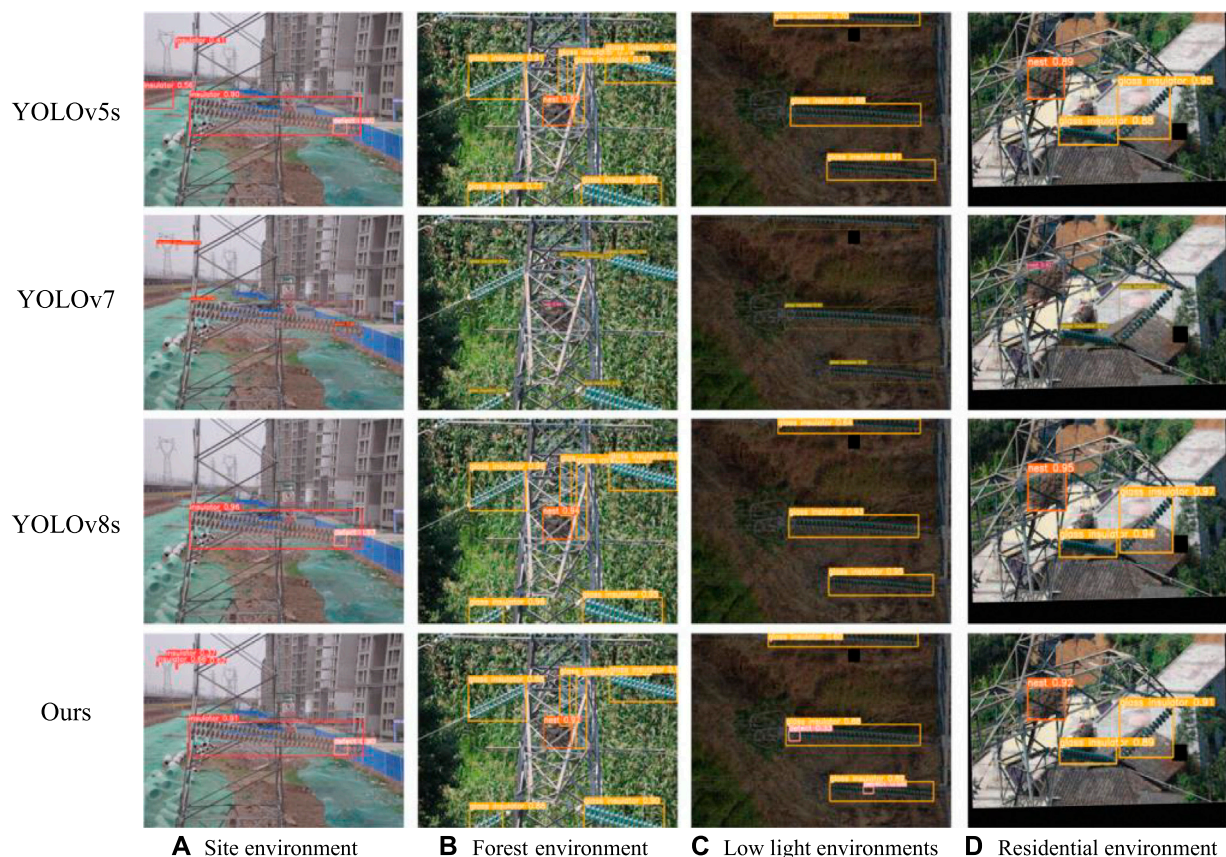


FIGURE 11
Comparison of detection results of different algorithms.

Finally, this paper compares the detection result graphs of the four models with the highest mAP values, and the comparison results are shown in Figure 11. From Figures 11B,D, it can be seen that the YOLOv7 and YOLOv8s algorithms have high detection accuracy for glass insulators, but the detection of the small targets as well as defective regions is not effective, and there are serious leakage cases.

From Figure C, it can be seen that YOLOv5s, YOLOv7 and YOLOv8s algorithms have lower detection accuracy under the interference of low-light as well as Gaussian noise, and the model's anti-interference ability is weaker, and the robustness is insufficient. In contrast, the FA-YOLOv5 proposed in this paper can accurately detect the small target insulators in the distance as well as the occluded insulators, and can accurately detect the insulators and their defective regions even under low light, and at the same time, it also has high detection accuracy under the interference of Gaussian noise and good anti-interference ability, which further proves that this paper's method can be applied to the presence of small targets under complex backgrounds in the presence of occlusion as well as dense and other cases can have a better detection effect.

5 Conclusion

In view of the slow detection speed and low accuracy, even leakage detection and false detection caused by the current insulator defect detection model with large number of parameters and large computation, as well as the complex environment in which insulators are located, the small percentage of defective parts, and the existence of mutual occlusion between insulators, this paper improves the YOLOv5s algorithm and proposes a lightweight FA-YOLOv5s algorithm based on it, with the following main contributions.

- 1) Strengthening feature fusion: By integrating the CBAM attention mechanism into the C3 module, the characteristics of both the attention mechanism and the C3 module are combined to enhance the algorithm's ability to fuse target feature information. This allows the feature information to better propagate to the detection head, resulting in improved detection accuracy.
- 2) Lightweight processing: Lightweight improvement is made to the convolutional modules in the main network of the model, balancing the relationship between network structure complexity and detection performance, so that the network reduces the number of parameters and computations at the cost of a small decrease in accuracy.
- 3) In this paper, the CIoU loss function used in the original model is improved to a WIoU loss function, which balances the variability in sample quality and improves the overlap between the prediction frame and the bounding box to improve the accuracy of the detection compared to CIoU.
- 4) A visualized software interface for defective insulator detection is designed, which enables a more intuitive observation of the detection results of the model.

However, during the experimental process, we found that the insulator defective dataset used in this paper is of a single type, and the data enhancement method can only expand the number of samples, and cannot enrich the diversity of the background environment, resulting in limited application in real scenarios. In the next work, we will consider going to the field to actually shoot more insulator images in different scenes, to further improve the robustness of the algorithm and the diversity of the dataset, and optimize the effect of YOLOv5s algorithm on the detection of defective insulators.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary materials, further inquiries can be directed to the corresponding author.

Author contributions

JL: Funding acquisition, Resources, Writing–review and editing. MH: Data curation, Methodology, Writing–original draft. JD: Investigation, Validation, Writing–review and editing. XL: Conceptualization, Resources, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China (62176067); Joint Fund for Basic and Applied Basic Research in Guangdong Province (20A1515111162); Scientific and Technological Planning Project of Guangzhou (201903010041, 202103000040); Key Project of Guangdong Province Basic Research Foundation (2020B1515120095); Project Supported by Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2019).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). *Yolov4: optimal speed and accuracy of object detection*. arXiv preprint arXiv:2004.10934, Available at: <https://doi.org/10.48550/arXiv.2004.10934>.
- Chen, B. (2020). Fault statistics and analysis of 220-kv and above transmission lines in A southern coastal provincial power Grid of China. *IEEE Open Access J. Power Energy* 7, 122–129. doi:10.1109/oajpe.2020.2975665
- Chen, J., Kao, S.-h., He, H., Zhuo, W., Wen, S., Lee, C.-H., et al. (2023). Run, don't walk: chasing higher flops for faster neural networks. arXiv preprint arXiv:2303.03667, Available at: <https://doi.org/10.48550/arXiv.2303.03667>.
- Chen, Y., Liu, H., Chen, J., Hu, J., and Zheng, E. (2023). Insu-YOLO: an insulator defect detection algorithm based on multiscale feature fusion. *Electronics* 12 (15), 3210. doi:10.3390/electronics12153210
- El-Hag, A. (2021). Application of machine learning in outdoor insulators condition monitoring and diagnostics. *IEEE Instrum. Meas. Mag.* 24, 101–108. doi:10.1109/mim.2021.9400959
- Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: a retrospective. *Int. J. Comput. Vis.* 111, 98–136. doi:10.1007/s11263-014-0733-5
- Gao, Z., Yang, G., Li, E., and Liang, Z. (2021). Novel feature fusion module-based detector for small insulator defect detection. *IEEE Sensors J.* 21, 16807–16814. doi:10.1109/jsen.2021.3073422
- Han, G., He, M., Gao, M., Yu, J., Liu, K., and Qin, L. (2022). Insulator breakage detection based on improved YOLOv5. *Sustainability* 14, 6066. doi:10.3390/su14106066
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). “Mask R-CNN,” in Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22–29 October 2017, 2961–2969.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. pattern analysis Mach. Intell.* 37, 1904–1916. doi:10.1109/tpami.2015.2389824
- Lewy, D., and Mańdziuk, J. (2023). An overview of mixing augmentation methods and augmentation strategies. *Artif. Intell. Rev.* 56, 2111–2169. doi:10.1007/s10462-022-10227-z
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). “Feature pyramid networks for object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2117–2125.
- Lin, T. Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., et al. (2014). “Microsoft coco: common objects in context,” in Proceedings, Part V 13 Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, 740–755. Springer.
- Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). “Path aggregation network for instance segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 8759–8768.
- Lu, H., Li, Y., Mu, S., Wang, D., Kim, H., and Serikawa, S. (2017). Motor anomaly detection for unmanned aerial vehicles using reinforcement learning. *IEEE internet things J.* 5, 2315–2322. doi:10.1109/ijot.2017.2737479
- Miao, X., Liu, X., Chen, J., Zhuang, S., Fan, J., and Jiang, H. (2019). Insulator detection in aerial images for transmission line inspection using single shot multibox detector. *IEEE Access* 7, 9945–9956. doi:10.1109/access.2019.2891123
- Quan, Y., Zhang, D., Zhang, L., and Tang, J. (2022). Centralized feature pyramid for object detection. arXiv preprint arXiv:2210.02093, Available at: <https://doi.org/10.48550/arXiv.2210.02093>.
- Raimundo, A. (2020). *Insulator data set-Chinese power line insulator dataset (cplid)*. IEEE Dataport.
- Redmon, J., and Farhadi, A. (2018). YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767, Available at: <https://doi.org/10.48550/arXiv.1804.02767>.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. *Adv. neural Inf. Process. Syst.* 28, 1137–1149. doi:10.1109/TPAMI.2016.2577031
- Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). “Generalized intersection over union: a metric and A loss for bounding box regression,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 658–666.
- Shuang, F., Wei, S., Li, Y., Gu, X., and Lu, Z. (2023). Detail R-CNN: insulator detection based on detail feature enhancement and metric learning. *IEEE Trans. Instrum. Meas.* 72, 1–14. Art no. 2524414. doi:10.1109/tim.2023.3305667
- Tan, P., Li, X. f., Ding, J., Cui, Z. s., Ma, J. e., Sun, Y. l., et al. (2022). Mask R-CNN and multifeature clustering model for catenary insulator recognition and defect detection. *J. Zhejiang University-SCIENCE A* 23, 745–756. doi:10.1631/jzus.a2100494
- Tong, Z., Chen, Y., Xu, Z., and Yu, R. (2023). Wise-iou: bounding box regression loss with dynamic focusing mechanism. arXiv preprint arXiv:2301.10051, Available at: <https://doi.org/10.48550/arXiv.2301.10051>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Advances in neural information processing systems* 30 (NeurIPS Proceedings).
- Wang, J., Zhang, T., Xue, X., and Chen, L. (2022). “Real-time recognition of transmission line insulators under complex backgrounds: a YOLOv5s approach,” in 2022 4th International Conference on Power and Energy Technology (ICPET), Beijing, China, 28–31 July 2022, 77–83.
- Wei, L., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). “Ssd: single shot multibox detector,” in Proceedings, Part I 14. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016. Springer International Publishing.
- Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). “Cbam: convolutional Block attention module,” in Proceedings of the European conference on computer vision (ECCV), 3–19.
- Yang, W., Liu, Y., and Li, Y. (2021). “Insulator UV image fault detection based on deep learning,” in 2021 IEEE International Conference on Electrical Engineering and Mechatronics Technology (ICEEMT) (IEEE), Qingdao, China, 02–04 July 2021, 632–635.
- Yang, Z., Xu, Z., and Wang, Y. (2022). Bidirection-Fusion-YOLOv3: an improved method for insulator defect detection using uav image. *IEEE Trans. Instrum. Meas.* 71, 1–8. doi:10.1109/tim.2022.3201499
- Yi, W., Ma, S., and Li, R. (2023). Insulator and defect detection model based on improved yolo-S. *IEEE Access* 11, 93215–93226. doi:10.1109/access.2023.3309693
- Yu, J., Jiang, Y., Wang, Z., Cao, Z., and Huang, T. (2016). “Unitbox: an advanced object detection network,” in Proceedings of the 24th ACM international conference on Multimedia, Amsterdam, The Netherlands, October 15 - 19, 2016, 516–520.
- Yu, Y., Cao, H., Wang, Z., Li, Y., Li, K., and Xie, S. (2019). Texture-and-shape based active contour model for insulator segmentation. *IEEE Access* 7, 78706–78714. doi:10.1109/access.2019.2922257
- Zhang, L.-P., Zhao, J.-M., and Ren, Y.-F. (2018). “Research on multiple features extraction technology of insulator images,” in 2018 10th International Conference on Modelling, Identification and Control (ICMIC) (IEEE), Guiyang, China, 02–04 July 2018, 1–6.
- Zhang, S., Qu, C., Ru, C., Wang, X., and Li, Z. (2023). Multi-objects recognition and self-explosion defect detection method for insulators based on lightweight GhostNet-YOLOV4 model deployed onboard UAV. *IEEE Access* 11, 39713–39725. doi:10.1109/access.2023.3268708
- Zhang, Y.-F., Ren, W., Zhang, Z., Jia, Z., Wang, L., and Tan, T. (2022). Focal and efficient iou loss for accurate bounding box regression. *Neurocomputing* 506, 146–157. doi:10.1016/j.neucom.2022.07.042
- Zhao, W., Xu, M., Cheng, X., and Zhao, Z. (2021). An insulator in transmission lines recognition and fault detection model based on improved faster RCNN. *IEEE Trans. Instrum. Meas.* 70, 1–8. Art no. 5016408. doi:10.1109/tim.2021.3112227
- Zheng, H., Sun, Y., Liu, X., Djike, C. L. T., Li, J., Liu, Y., et al. (2020a). Infrared image detection of substation insulators using an improved fusion single shot multibox detector. *IEEE Trans. Power Deliv.* 36, 3351–3359. doi:10.1109/tpwrd.2020.3038880
- Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., and Ren, D. (2020b). Distance-iou loss: faster and better learning for bounding box regression. *Proc. AAAI Conf. Artif. Intell.* 34, 12993–13000. doi:10.1609/aaai.v34i07.6999
- Zheng, Z., Wang, P., Ren, D., Liu, W., Ye, R., Hu, Q., et al. (2021). Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* 52, 8574–8586. doi:10.1109/tcyb.2021.3095305



OPEN ACCESS

EDITED BY

Haris M. Khalid,
University of Dubai, United Arab Emirates

REVIEWED BY

Mohammad Amir,
Indian Institutes of Technology (IIT), India
Minh Quan Duong,
The University of Danang, Vietnam
Shuai Lu,
Southeast University, China

*CORRESPONDENCE

Haodong Du,
✉ 3170421159@stu.xaut.edu.cn

RECEIVED 03 July 2023

ACCEPTED 24 October 2023

PUBLISHED 03 November 2023

CITATION

Zhang H, Jia R, Du H, Liang Y and Li J
(2023), Short-term interval prediction of
PV power based on quantile regression-
stacking model and tree-structured
parzen estimator optimization algorithm.
Front. Energy Res. 11:1252057.
doi: 10.3389/fenrg.2023.1252057

COPYRIGHT

© 2023 Zhang, Jia, Du, Liang and Li. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Short-term interval prediction of PV power based on quantile regression-stacking model and tree-structured parzen estimator optimization algorithm

Hongyang Zhang^{1,2}, Rong Jia¹, Haodong Du^{1*}, Yan Liang¹ and Jiangfeng Li¹

¹School of Electrical Engineering, Xi'an University of Technology, Xi'an, China, ²Shaanxi Province Distributed Energy Co., Xi'an, China

In recent years, the photovoltaic (PV) industry has grown rapidly and the scale of grid-connected PV continues to increase. The random and fluctuating nature of PV power output is beginning to threaten the safe and stable operation of the power system. PV power interval forecasting can provide more comprehensive information to power system decision makers and help to achieve risk control and risk decision. PV power interval forecasting is of great importance to power systems. Therefore, in this study, a Quantile Regression-Stacking (QR-Stacking) model is proposed to implement PV power interval prediction. This integrated model uses three models, extreme gradient boosting (Xgboost), light gradient boosting machine (LightGBM) and categorical boosting (CatBoost), as the base learners and Quantile Regression-Long and Short Term Memory (QR-LSTM) model as the meta-learner. It is worth noting that in order to determine the hyperparameters of the three base learners and one meta-learner, the optimal hyperparameters of the model are searched using a Tree-structured Parzen Estimator (TPE) optimization algorithm based on Bayesian ideas. Meanwhile, the correlation coefficient is applied to determine the input characteristics of the model. Finally, the validity of the proposed model is verified using the actual data of a PV plant in China.

KEYWORDS

photovoltaic Forecast, interval Forecast, Optimization, Stacking, Photovoltaic

1 Introduction

In recent years, the human demand for electrical energy has been increasing. At present, thermal power generation occupies 60% of the global electricity energy supply, however, thermal power generation requires a large amount of non-renewable energy in the production process, and the non-renewable energy sources stored on the Earth, such as coal, oil and natural gas, are becoming increasingly depleted (Viet et al., 2020), and the energy crisis has sounded an alarm for mankind (Frilingou et al., 2023). Therefore, accelerate the energy revolution, optimize the energy structure is urgent to achieve sustainable development of energy has become a key concern of countries around the world. Solar energy is a renewable energy source with great potential, and countries around the world have reached a consensus on the need for solar energy development, of which

photovoltaic power generation is an important way of solar energy development and utilization (Rafique et al., 2020; Khalid et al., 2023). With the progress of technology and cost reduction, photovoltaic power generation has been widely promoted and applied in all countries around the world, and the installed capacity has been rising in recent years.

PV power output is uncontrollable and subject to various meteorological factors, showing strong volatility, randomness, intermittency and non-smoothness. The PV power system is equivalent to an uncontrollable power source for the power system. With the increasing scale of grid-connected PV, unstable PV power output will cause difficulties in power system scheduling and real-time power balancing. At the same time, PV power output fluctuations can lead to sharp fluctuations in frequency and voltage, and the resulting shocks may threaten the safe and stable operation of the power system. In addition, large-scale grid-connected PV may have a certain negative impact on the damping characteristics of the power system, which in turn threatens the safe and stable operation of the power system (Rafique et al., 2022). Therefore, in order to improve the security of the power system and the reliability of power supply, the light has to be abandoned. Accurate PV power prediction helps the power system scheduling department to reasonably arrange the power system scheduling plan and realise the real-time balance of power generation and power consumption, so as to ensure that the power system can operate reliably, safely and stably. For PV power operating companies, it can improve the economic efficiency of PV power plants. In addition to this, energy storage technology has a very high potential in reducing the threat of PV fluctuations to the power system (Amir et al., 2023).

According to the different mechanisms within the prediction models, PV output forecasting models can be categorized into: physical models, statistical models, machine learning models and integrated models. The physical prediction model uses the installation position, tilt angle, design parameters, operating characteristics and conversion efficiency of PV modules to establish a physical model, while meteorological data such as solar irradiance is used as the data basis for the physical model to obtain the predicted value of PV power generation through the calculation of the physical model (Dolara et al., 2015). The statistical model is only data-driven. The statistical model inputs weather variables such as solar irradiance and historical data of PV power, and extracts the intrinsic correlation information to build a mapping model to achieve the prediction of future PV output (Gellert et al., 2022). While traditional statistical methods have very limited nonlinear modeling capability, machine learning prediction models (Rao et al., 2022) have powerful nonlinear mapping modeling capability, which has led to its rapid development in the field of PV forecasting. Twenty-four machine learning models were developed for implementing PV power prediction by Dávid Markovics et al. Day-ahead PV power prediction was performed based on numerical weather forecast data. The effects of predictor variable selection and the benefits of hyperparameter tuning were also investigated in detail in this study (Markovics and Mayer, 2022). In recent years many researchers have turned to the development and research work of combined prediction models (Liang et al., 2023), which have superior predictive performance, model generalization performance, and robustness.

Traditional PV power prediction techniques focus on point prediction. The output of point prediction is a single point expected value of PV power at a certain moment in the future. However, due to the chaotic nature of the atmospheric system, PV power prediction errors cannot be avoided. There are significant uncertainties in the prediction results, and the information provided by point prediction is very limited. In contrast to point prediction, interval prediction of PV power uses prediction intervals to achieve a quantitative estimate of the prediction uncertainty. The interval prediction results provide the upper and lower bounds of the fluctuation of PV power at a certain confidence level at a certain time in the future, which makes up for the limitations of the PV power point prediction technique and can provide more comprehensive data support for the power system. The PV power interval prediction results can provide important references for the operational risk assessment and risk decision-making of the power system, and further improve the security and economy of the power system. In addition, PV power interval prediction technology has a very broad application prospect in the fields of power system planning, power system scheduling, energy storage configuration and regulation, and power market trading. Zhenhao Wang et al. (Wang et al., 2022) established a deep convolutional generative adversarial network model to generate PV power characteristic curves in different scenarios, and then established a QRLSTM model to achieve PV power interval prediction. Ming Ma et al. (Ma et al., 2022) analysed the distribution of PV power prediction errors and then constructed PV power prediction intervals using a kernel density estimation algorithm.

The existing PV power interval prediction is mainly realised using a single model, and its prediction performance needs to be further improved. Multi-model fusion technology will be an important development direction in the future. The values of the model parameters largely determine the prediction performance of the model, so the hyperparameter optimisation problem of the fusion model needs further research. Therefore, in this paper a stacking model that can achieve the prediction of PV power intervals is proposed. An optimisation algorithm is used to determine the optimal hyperparameter values for this model to improve the PV power interval accuracy. In order to provide important data support for power system operation risk assessment and risk decision-making, and to further improve the safety and economy of the power system. The main contributions of this paper are as follows:

- 1) A novel QR-Stacking integrated model is proposed to implement PV power interval prediction. Multiple decision tree models are used as the base learners of this integrated model, and deep neural networks are used as the meta-learner of this integrated model. The QR-Stacking integrated model is constructed by combining the quantile regression model and the Stacking integrated model to achieve the PV power interval prediction. This is the first application of this stacking model in the field of PV power interval prediction.
- 2) To improve the prediction accuracy of the QR-Stacking integrated model, the Tree-structured Parzen Estimator algorithm was used to search for determining the hyperparameters of multiple base learners and a meta-learner.

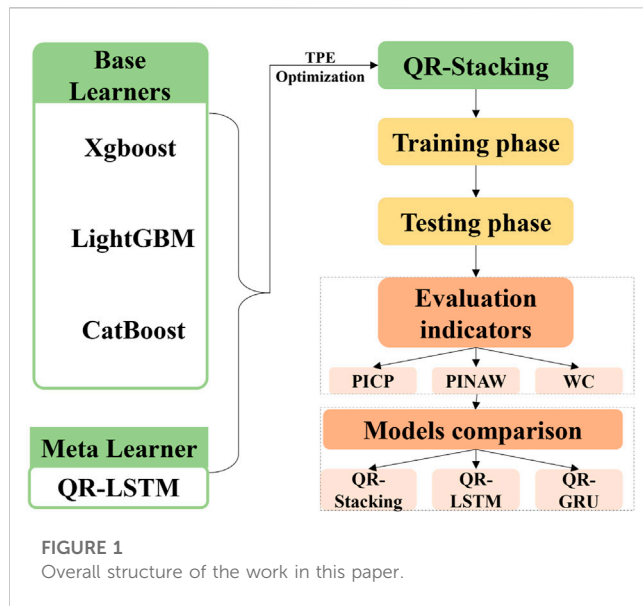


TABLE 1 Nomenclature.

Abbreviations	Full text
CatBoost	categorical boosting
DI	direct radiation
EI	expected improvement
GRU	Gated Recurrent Unit
GI	global irradiance
H	humidity
Ken	Kendall correlation coefficient
LSTM	Long and Short Term Memory
LightGBM	light gradient boosting machine
PV	photovoltaic
PICP	Prediction interval coverage probability
PINAW	Prediction interval normalized average width
QR	Quantile Regression
QR-Stacking	Quantile Regression-Stacking
QR-LSTM	Quantile Regression-Long and Short Term Memory
QR-GRU	Quantile Regression-Gated Recurrent Unit
TPE	Tree-structured Parzen Estimator
T	temperature
WC	comprehensive evaluation index
WD	wind direction
Xgboost	extreme gradient boosting

- 3) Simulation analysis was conducted based on actual PV power generation data from a PV power plant in China. Multiple interval evaluation metrics were used to evaluate the prediction intervals. A

comprehensive comparative analysis of the interval prediction accuracy of QR-Stacking and multiple benchmark prediction models was performed. The results show that the model can give full play to the advantages of each algorithm and improve the prediction accuracy of PV power intervals.

The rest of the paper is organized as follows: Section 2 describes the prediction models and optimization algorithms used in this paper. Section 3 describes in detail the evaluation metrics of the prediction model. Section 4 provides case studies. The prediction performance of the proposed prediction model is compared with that of several benchmark models. The accuracy of the prediction models is verified by experimental simulations. Section 5 summarizes the whole paper.

2 Methodology

Figure 1 shows the overall structure of the work in this paper. Firstly, three base learners, Xgboost, LightGBM and CatBoost, and a meta-learner, QR-LSTM, are built. The proposed QR-Stacking model is constructed from the above four models and the TPE optimisation algorithm. Secondly, the proposed model is trained and tested using real PV data. Finally, three evaluation metrics are used to compare and analyse the prediction performance of the proposed model with QR-LSTM and QR-GRU models. The nomenclature used in this paper is presented in Table 1.

2.1 Stacking

It has been shown that single prediction models have limited prediction accuracy. Ensemble machine learning would be an important solution to this challenge. Usually, the first layer of the stacking model is the base learner layer and the second layer is the meta-learner layer. The meta-learner layer corrects the prediction error of the base learner. In this research, Xgboost, LightGBM and CatBoost are used as base learners. The QR-LSTM model is used as a meta-learner. The following section details the modeling principles of the three base learners and one meta-learner.

2.2 Base learners

2.2.1 Xgboost

The XGBoost algorithm is an improved algorithm of the gradient augmented regression tree. The main improvements of the XGBoost algorithm are the improvement of the objective function and its solving method.

The objective function (loss function) of the XGBoost algorithm during training consists of two parts, as shown in Eq. 1.

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

Where $\sum_{i=1}^n l(y_i, \hat{y}_i)$ is used to characterize the difference between the predicted values \hat{y}_i and true values y_i . $\sum_{k=1}^K \Omega(f_k)$ is the regularization term.

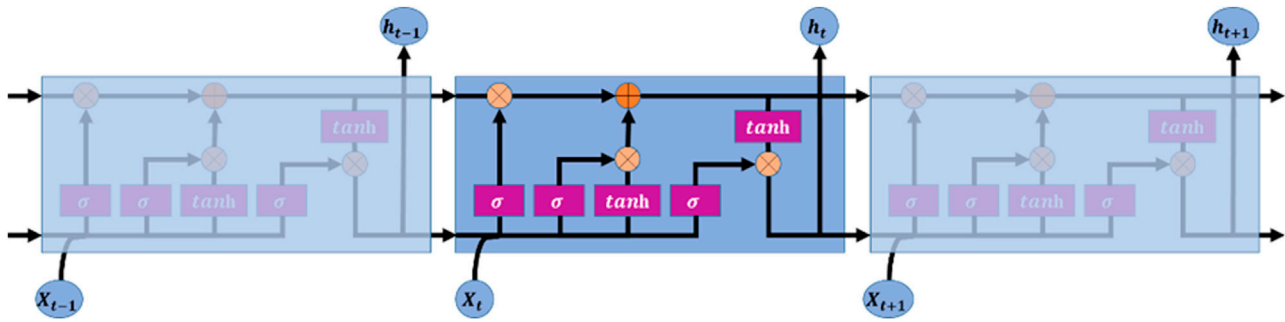


FIGURE 2
The structure of the LSTM unit.

Algorithm 1 Tree-structured Parzen Estimator optimization algorithm

Require: Total number of initialised sampling points m , Objective function $f(s)$, Hyperparametric search space S , Agent model TPE , Acquisition function (Expected Improvement) EI

Ensure: Historical sampling set $D(s, f)$, Optimal hyperparameters s^* and its corresponding objective function value

- 1: Choose n_0 sampling points in the hyperparameter space S and calculate the value of $f(s)$ at these sampling points
- 2: Update the sampling data set $D(s, f(s))$
- 3: $n = n_0$
- 4: **for all** $n = n_0 + 1, n_0 + 2, \dots, m$ **do**
- 5: Fitting the agent model TPE with the current dataset $D(s, f(s))$ enables model updating.
- 6: Select the superparameter combination s that maximises acquisition function EI .
- 7: Compute the value of the function $f(s)$ when the hyperparameter is configured as s .
- 8: Add the new $(s, f(s))$ to the existing dataset D
- 9: **end for**
- 10: **return** Historical sampling set $D(s, f(s))$, Optimal hyperparameters s^* and its corresponding objective function value

FIGURE 3
Pseudo-code of the TPE optimisation algorithm.

The regularization term can be calculated from Eq. 2:

$$\sum_{k=1}^K \Omega(f_k) = \gamma T + \frac{1}{2} \theta \sum_{j=1}^T \omega_j^2 \quad (2)$$

where $\Omega(f_k)$ is a function of the complexity of the decision tree f_k . γ is the penalty term of the L1 regular. T is the total number of leaf nodes of the decision tree. θ is the penalty term of the L2 regular. ω_j is the weight of the decision tree f_i at the j th leaf node.

Each iteration updates the objective function to Eq. 3.

$$Obj^t = \sum_{i=1}^n l[y_i, \hat{y}_i^{(t-1)} + f_t(x_i)] + \Omega(f_t) \quad (3)$$

Using a second-order Taylor expansion for the above equation, the following equation is obtained by removing the constant term.

$$Obj^t = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i \times f_t^2(x_i) \right] + \Omega(f_t) \quad (4)$$

where $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$ are the first- and second-order derivatives of the objective function, respectively.

2.2.2 LightGBM

The basic idea of LightGBM is to obtain the final strong regression tree using multiple iterations of the weak regression tree. The new regression tree obtained from each iteration is obtained by fitting the prediction residuals of the previous regression tree. Finally, the outputs of all regression trees are summed to output the better-performing results. The calculation is shown in Eq. 5.

$$F(x) = \sum_{m=1}^M f_m(x) \quad (5)$$

where $f_m(x)$ is the output value of the m th weak regression tree and $F(x)$ is the final output value of the model.

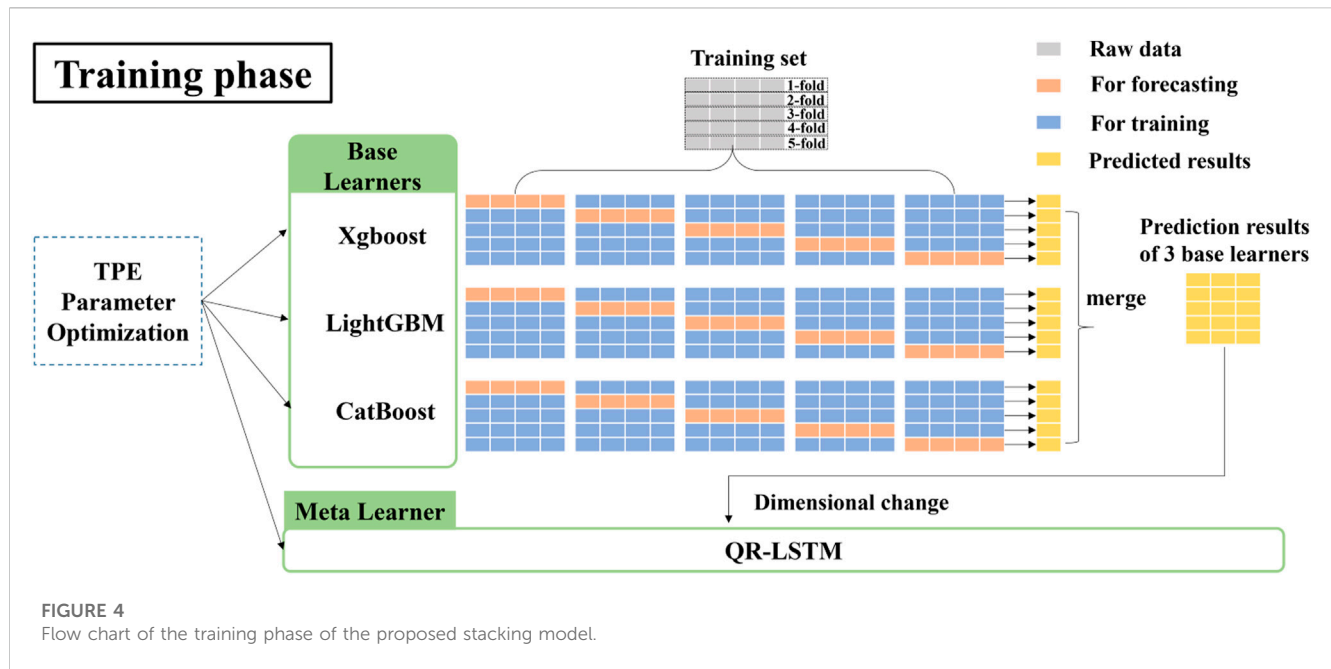


FIGURE 4
Flow chart of the training phase of the proposed stacking model.

Algorithm 2 QR-Stacking Training phase

Require: Training dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, Basic Learner Algorithms $B = \{Xgboost, LightGBM, CatBoost, \dots\}$, Meta-Learner Algorithm $M = \{QR - LSTM\}$, Optimal hyperparameters for basic learners and the meta-learner determined by the TPE algorithm $H = \{H_{Xgboost}, H_{LightGBM}, H_{CatBoost}, H_{QR-LSTM}\}$

Ensure: Stacking predictive model $QR - Stacking$

```

for all base model  $B_i$  do
  2: for all fold = 1, 2, ..., 5 do
    4/5 of  $D$  was used to train  $B_i[fold]$ 
  4: Using the trained model  $B_i[fold]$  to predict the remain 1/5 of  $D$  to get  $P_{B_i}[fold]$ 
  end for
  6: Stack  $P_{B_i}[1], P_{B_i}[2], \dots, P_{B_i}[5]$  to get  $P_{B_i}$ 
end for
8: Train meta-level model  $M$  on  $P_B = \{P_{B_1}, P_{B_2}, P_{B_3}\}$  and  $D_y = \{y_1, y_2, \dots, y_n\}$ 
 $QR - Stacking \leftarrow$  Ensemble of base models  $B_1, B_2, B_3$  and  $M$ 
10: return  $QR - Stacking$ 

```

FIGURE 5
Pseudo-code for the training phase of the proposed model.

2.2.3 CatBoost

The CatBoost model is an improved gradient boosted decision tree (GBDT) model. The improvements of CatBoost over traditional GBDT are as follows:

Traditional GBDT derives the gradient of the current model based on the same dataset in each iteration of training, but this leads to biased point-by-point gradient estimation. CatBoost uses Ordered Boosting to improve the gradient estimation method of the traditional algorithm. The improved algorithm obtains an unbiased estimate of the gradient to mitigate the effect of the gradient estimation bias and thus improve the generalization ability of the model. To obtain unbiased gradient estimation, the

CatBoost model trains a separate model M_i for each sample x_i , which is obtained by training with a training set that does not contain sample x_i . Then M_i is used to obtain a gradient estimate on the samples. Finally, this gradient is used to train the weak learner and obtain the final model.

2.3 Meta-learner

2.3.1 Quantile regression

The quantile regression (QR) model is used to study the relationship between the conditional quantiles of the independent

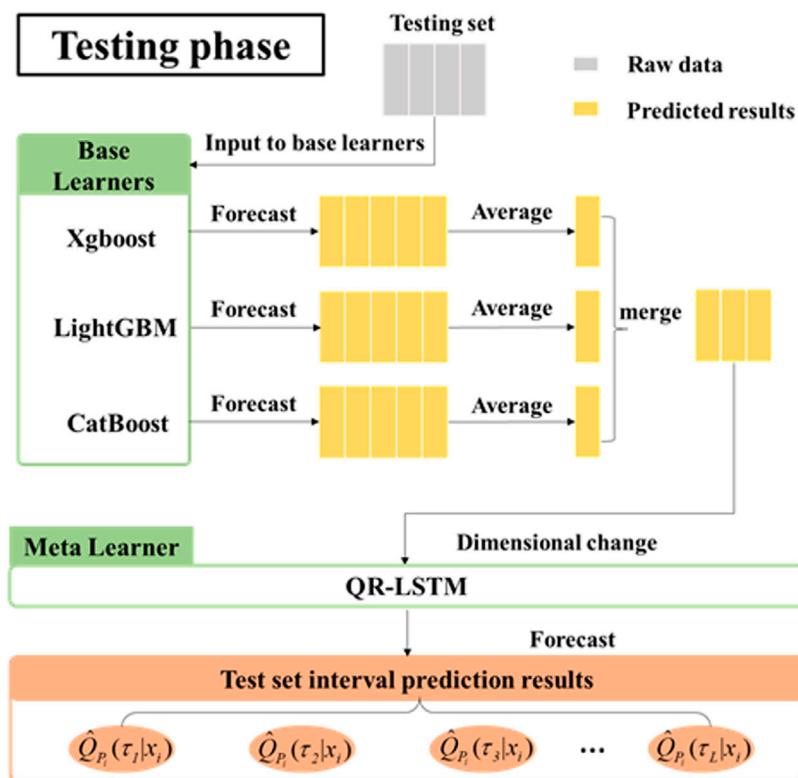


FIGURE 6
Flow chart of the testing phase of the proposed stacking model.

Algorithm 3 QR-Stacking testing phase

Require: Stacking predictive model *QR-Stacking*, test dataset $D_{test} = \{x_1, x_2, \dots, x_m\}$
Ensure: Predictions P_{test} for D_{test}
for all base model B_i in B **do**
 for all fold = 1, 2, ..., 5 **do**
 3: Using the trained model $B_i[fold]$ generate predictions $T_{Bi}[fold]$ for D_{test}
 end for
 Average $T_{Bi}[1], T_{Bi}[2], \dots, T_{Bi}[5]$ according to the row to get T_{Bi}
6: **end for**
 $P_{test} \leftarrow$ Predictions of M on $T_B = \{T_{B_1}, T_{B_2}, T_{B_3}\}$
return P_{test}

FIGURE 7
Pseudo-code for the testing phase of the proposed model.

and dependent variables. The quantile regression model can be represented by Eq. 6.

$$Q_{P_i}(\tau|x_i) = x_i\beta(\tau) \quad i = 1, 2, \dots, n \quad (6)$$

where $Q_{P_i}(\tau|x_i)$ is the conditional quantile. $\tau \in (0, 1)$. $\beta(\tau)$ is the vector of regression coefficients. $\beta(\tau) = [\beta_0(\tau), \beta_1(\tau), \dots, \beta_m(\tau)]^T$. Each element $\beta_j(\tau)$ in $\beta(\tau)$ characterizes the degree of influence of the j th independent variable on the dependent variable. n is the total

number of samples. P_i is the dependent variable. x_i is the independent variable. The dependent variable is usually multiple, i.e., $x_i = [x_{i,0}, x_{i,1}, \dots, x_{i,m}]$

The objective of solving the quantile regression model is $\beta(\tau)$. The problem can be solved by minimizing the loss function as shown in Eq. 7.

$$L = \sum_{i=1}^n \gamma_{\tau}(P_i - x_i\beta(\tau)) \quad (7)$$

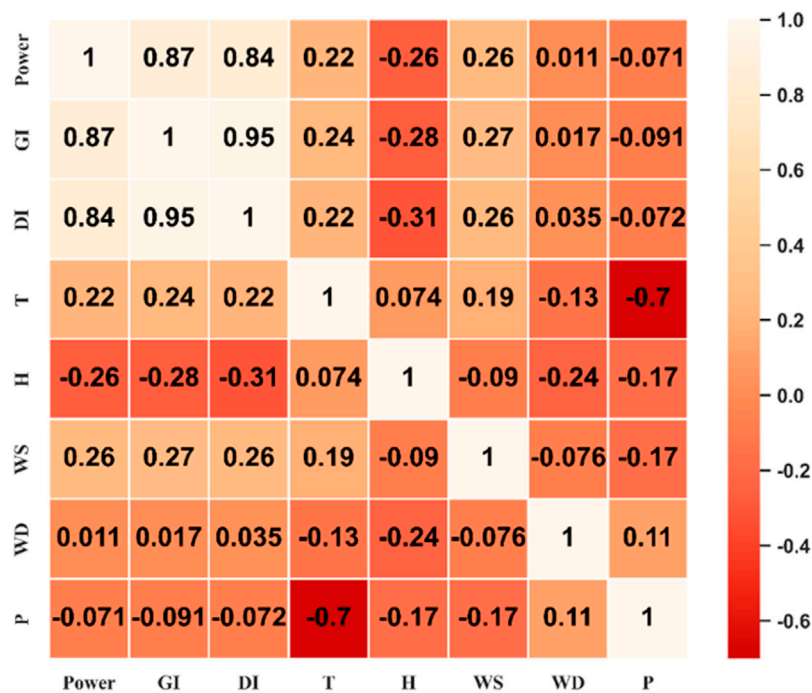


FIGURE 8
Heat map of correlation between power variables and meteorological factor variables.

TABLE 2 The hyperparameter search range settings for the base learners and meta-learner.

Model	Hyperparameter	Search range
Xgboost	learning_rate	Choice [0.01,0.03,0.1,0.2,0.5]
	n_estimators	Randint (100,1000)
	max_depth	Choice [4,6,8,10,12,15]
	min_child_weight	Randint (3,20)
LightGBM	learning_rate	Choice [0.01,0.03,0.1,0.2,0.5]
	n_estimators	Randint (100, 1000)
	max_depth	Choice [4,6,8,10]
	min_child_samples	Randint (0,30)
	min_child_weight	Randint (3,20)
CatBoost	learning_rate	Choice [0.01,0.03,0.1,0.2,0.5]
	iterations	Randint (100,1000)
	depth	Choice [4,6,8,10,15]
LSTM	units	Qrandint (16, 512,16)
	dropout	Choice [0.01,0.2,0.5,0.8,0.9]
	activation	Choice ["linear","relu","elu"]

where γ_τ is an asymmetric function with respect to the quantile τ . $\hat{\beta}(\tau)$ is the estimated value of $\beta(\tau)$. Its specific expression is given by:

$$\gamma_\tau(s) = \begin{cases} \tau s & s \geq 0 \\ (\tau - 1)s & s < 0 \end{cases} \quad (8)$$

where the expression of s is $s = P_i - x_i\beta(\tau)$.

Thus the expression for solving $\hat{\beta}(\tau)$ is as follows:

$$\hat{\beta}(\tau) = \operatorname{argmin} \sum_{i=1}^n \gamma_\tau(P_i - x_i\beta(\tau)) \quad (9)$$

Ultimately, the estimates obtained by quantile regression model estimation at different conditional quantiles are as follows.

$$\hat{Q}_{P_i}(\tau|x_i) = x_i\hat{\beta}(\tau) \quad i = 1, 2, \dots, n \quad (10)$$

2.3.2 Long and Short Term Memory

The Long and Short Term Memory (LSTM) model was first proposed by Hochreiter and Schmidhuber in the 1990s as a solution to the issue of vanishing gradients in traditional RNNs. The incorporation of gating units, consisting of forgetting, input, and output gates, allows LSTMs to selectively retain or discard information within the cell state, enabling them to effectively capture and model long-term dependencies in sequential data. As a result, LSTMs have become a widely utilized tool in the field of deep learning. Figure 2 is a schematic diagram of the structure of the LSTM model.

2.3.3 Quantile Regression-Long and Short Term Memory

Quantile regression model in the form of loss function and LSTM model are fused to achieve PV power interval prediction. The

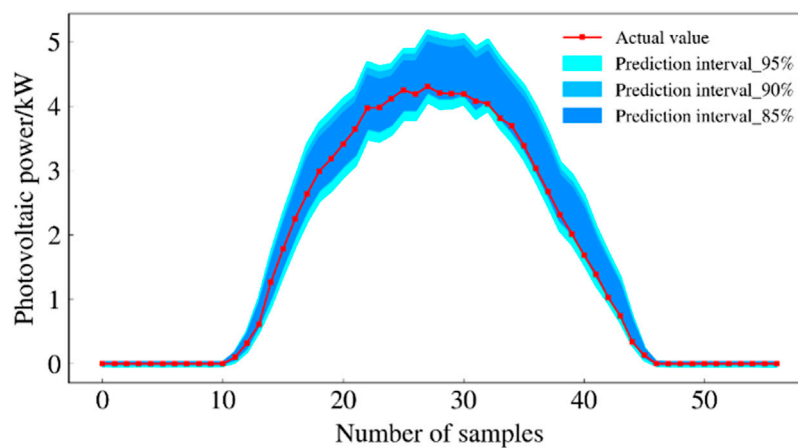


FIGURE 9

Forecast intervals of the QR-Stacking model during sunny days.

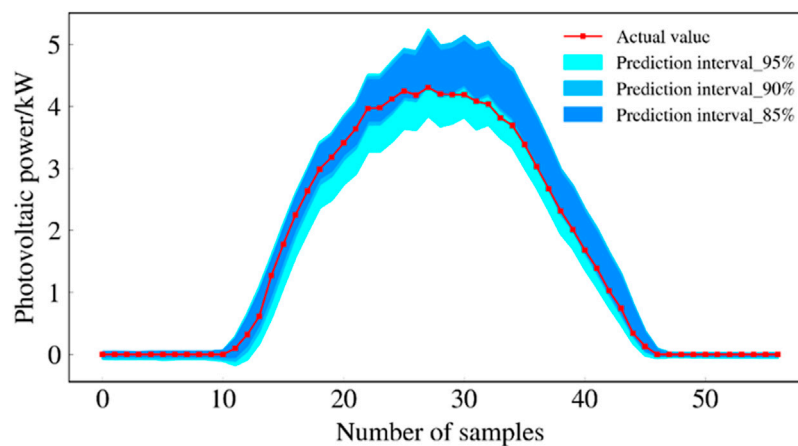


FIGURE 10

Forecast intervals of QR-LSTM model during sunny days.

Quantile Regression-Long and Short Term Memory (QR-LSTM) model serves as a meta-learner for the proposed model to further correct the prediction bias of the base learner.

2.4 Tree-structured parzen estimator

In this study, the Tree-structured Parzen Estimator (TPE) optimization algorithm is proposed to achieve the global optimization of each model hyperparameter. The TPE algorithm uses a probability density estimator based on the tree structure to implement Bayesian optimization. The TPE technique may fast converge to the global optimal solution and models the parameter space using a tree structure.

The main advantages of the TPE algorithm are (1) It avoids the inefficiencies of traditional grid search or random search by using probability density estimates to model the objective function. (Nguyen et al., 2020). (2) The TPE algorithm can automatically

adjust the direction and scope of the search. (3) The TPE algorithm can handle discrete, continuous, and mixed types of hyperparameters, making it applicable to a variety of machine learning models and algorithms. (4) The TPE algorithm is based on Bayesian optimisation theory, which has a solid mathematical foundation and reliable theoretical support. (5) The TPE algorithm estimates the probability density function in the parameter space by constructing a tree-like structure, which enables it to find high probability regions quickly and reduces the size of the search space. In contrast, optimisation algorithms such as genetic algorithms require a large number of iterations and crossover operations with high computational complexity. (6) The TPE algorithm is able to handle the noise in the objective function better and find the optimal solution more stably through the estimation of the probability density function. While optimisation algorithms such as genetic algorithm may be disturbed by noise and get unstable results.

The core of TPE optimisation is to find a set of hyperparameters that minimise the established objective function. The Bayesian-

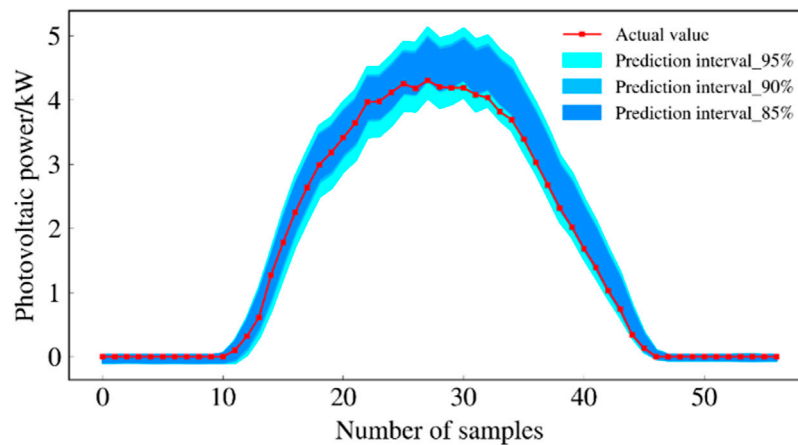


FIGURE 11

Forecast intervals of QR-GRU model during sunny days.

TABLE 3 Evaluation of the prediction results of each model during sunny days.

Model	Confidence levels (%)	PICP	PINAW	WC
QR-Stacking	95	1.000000	0.142233	0.142233
	90	0.912281	0.108516	0.118951
	85	0.859649	0.095189	0.110731
QR-LSTM	95	1.000000	0.159846	0.159846
	90	0.842105	0.107182	0.127279
	85	0.789473	0.090057	0.114073
QR-GRU	95	1.000000	0.146956	0.146956
	90	0.824561	0.105388	0.127812
	85	0.754385	0.093575	0.124041

based TPE optimisation algorithm reduces the number of computations and time cost by selecting the most promising set of hyperparameters for the next evaluation. Figure 3 illustrates the pseudo-code of the TPE optimisation algorithm. The following section describes in detail the selection criteria for the objective function and the next set of hyperparameters:

The goal of hyperparameter optimization is to find the value of the hyperparameter that minimizes the loss of the machine learning model. It can be expressed as Eq. 11.

$$s^* = \underset{s \in S}{\operatorname{argmin}} f(s) \quad (11)$$

where S is the optional hyperparameter space and s^* is the best set of hyperparameters.

The whole concept of Bayesian optimization is to reduce the number of computations and time cost by selecting the most promising set of hyperparameters as possible for the next evaluation. The selection criteria for the next set of hyperparameters is the expected improvement (EI), which is expressed as:

$$EI_{t^*}(s) = \int_{-\infty}^t (t^* - t) p(t, s) dt \quad (12)$$

where t^* is the threshold of the objective function, s is the proposed hyperparameter, t is the actual value of the objective function when the proposed hyperparameter s is used, and $p(t, s)$ denotes the agent probability model. $p(t, s)$ is defined in the TPE method, and $p(t, s)$ is denoted as

$$p(t, s) = \begin{cases} l(s) & \text{if } t < t^* \\ g(s) & \text{if } t > t^* \end{cases} \quad (13)$$

where $l(s)$ denotes the probability of hyperparameter set s when the value of the objective function t is less than a threshold value t^* , $g(s)$ denotes the probability of hyperparameter set s when the value of the objective function t is greater than a threshold value t^* .

The EI criteria when using the TPE method can be expressed as follows.

$$EI_{t^*}(s) = \int_{-\infty}^t (t^* - t) p(t, s) dt = \int_{-\infty}^t (t^* - t) \frac{p(s, t) p(t)}{p(s)} dt \quad (14)$$

$p(s)$ can be denoted as $p(s) = \int_{\mathcal{R}} p(s|t) p(t) dt = gl(s) + (1 - g)g(s)$. Let $\gamma = p(t < t^*)$. The final EI can be expressed as follows.

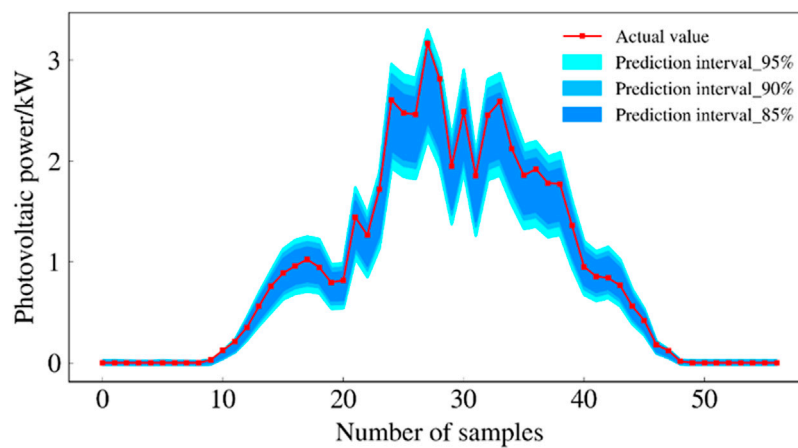


FIGURE 12

Forecast intervals of QR-Stacking model during rainy days.

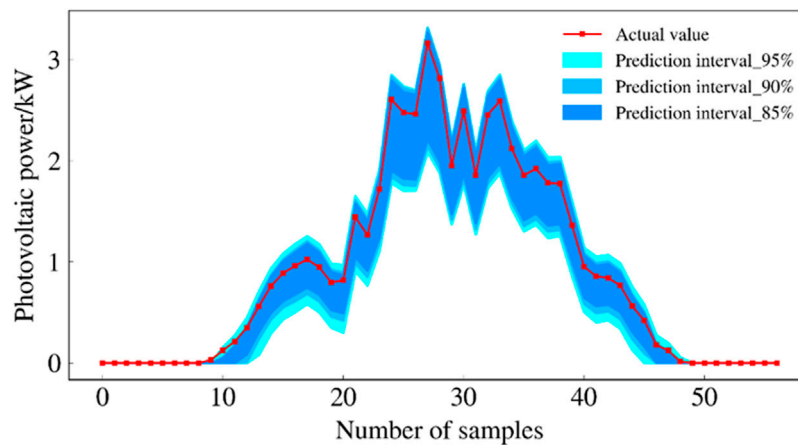


FIGURE 13

Forecast intervals of QR-LSTM model during rainy days.

$$EI_t^*(s) = \frac{\gamma t^* l(s) - l(s) \int_{-\infty}^t p(t) dt}{\gamma l(s) + (1 - \gamma) g(s)} \propto \left(\gamma + \frac{g(s)}{l(s)} (1 - \gamma) \right)^{-1} \quad (15)$$

In order to maximize EI, the ratio $\frac{l(s)}{g(s)}$ should be maximized. Therefore, the expected set of hyperparameters s has a higher probability under $l(s)$.

Eventually, through continuous iteration, the set of hyperparameters that can make the objective function achieve the minimum value is obtained. This set of hyperparameters is the best hyperparameters for the proposed model.

2.5 Quantile regression-stacking model optimized using the tree-structured parzen estimator algorithm for photovoltaic power interval prediction

In this study a stacking model using an efficient hyperparametric optimization method for PV power interval prediction is proposed.

Xgboost, LightGBM and CatBoost are used as the base learners. QR-LSTM is used as a meta-learner. Firstly, three basic learners are used to independently make predictions of PV power output, which are able to learn the trend of PV power from historical data. Each basic learner produces a set of predictions. Then, the prediction results of these base learners are fed into QR-LSTM to achieve the final prediction. QR-LSTM further corrects the prediction errors of the three base learners to improve the prediction accuracy. Notably, the quantile regression model in the QR-LSTM model is capable of constructing prediction intervals to quantify the uncertainty in PV power prediction. By combining the strengths of these learners, the QR-Stacking model is able to better address the challenges associated with PV power output fluctuations and more accurately quantify the uncertainty in PV power forecasts. In addition to this, the TPE algorithm is also used to search for the optimal parameters of the base and meta learners to further improve the interval prediction performance of the model.

The proposed stacking model is illustrated separately in a training phase and a testing phase.

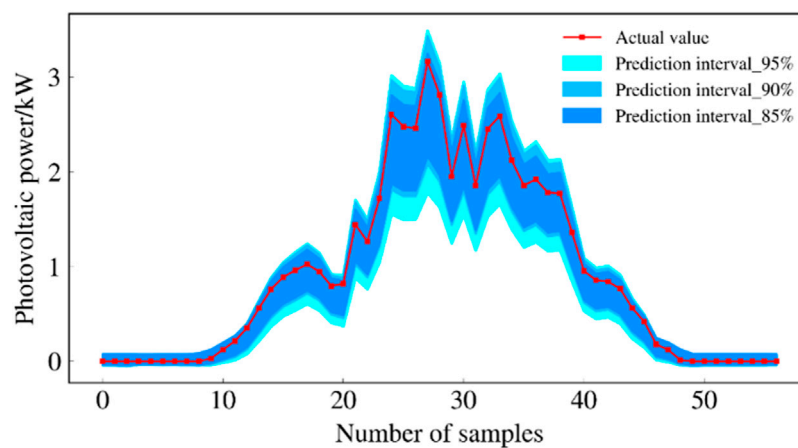


FIGURE 14

Forecast intervals of QR-GRU model during rainy days.

TABLE 4 Evaluation of prediction results of each model during cloudy and rainy days.

Model	Confidence levels (%)	PICP	PINAW	WC
QR-Stacking	95	0.982456	0.135335	0.137751
	90	0.964912	0.110254	0.114264
	85	0.894736	0.084793	0.094768
QR-LSTM	95	0.982456	0.152176	0.154893
	90	0.912280	0.127719	0.140000
	85	0.912280	0.110009	0.120587
QR-GRU	95	1.000000	0.180607	0.180607
	90	1.000000	0.148593	0.148593
	85	1.000000	0.124890	0.124890

2.5.1 Training phase

The steps of the stacked model training phase are shown in Figure 4. The pseudo-code for the training phase of the proposed model is presented in Figure 5. The 5-fold cross-validation and TPE parameter optimization methods are used in the training phase of the proposed stacking model. The details are illustrated below.

- 1) The training set is divided into 5 folds.
- 2) Training the Xgboost model using a 5-fold cross-validation method and a TPE parameter optimization method. In the first iteration, the last 4 folds are used for training the model and the first fold is used for prediction. The obtained prediction result is $P_{Xgboost}^{1-fold}$. In the second iteration, the second fold is used for prediction and the remaining four folds are used for training. The obtained prediction result is $P_{Xgboost}^{2-fold}$. This process is repeated until the prediction results are obtained for all 5 folds. Finally, the prediction result obtained by the Xgboost model is:

$$P_{Xgboost}^{Train} = [P_{Xgboost}^{1-fold}, P_{Xgboost}^{2-fold}, P_{Xgboost}^{3-fold}, P_{Xgboost}^{4-fold}, P_{Xgboost}^{5-fold}] \quad (16)$$

- 3) Using the same process as (16), the outputs obtained from the LightGBM and CatBoost models are expressed as follows.

$$P_{LightGBM}^{Train} = [P_{LightGBM}^{1-fold}, P_{LightGBM}^{2-fold}, P_{LightGBM}^{3-fold}, P_{LightGBM}^{4-fold}, P_{LightGBM}^{5-fold}] \quad (17)$$

$$P_{CatBoost}^{Train} = [P_{CatBoost}^{1-fold}, P_{CatBoost}^{2-fold}, P_{CatBoost}^{3-fold}, P_{CatBoost}^{4-fold}, P_{CatBoost}^{5-fold}] \quad (18)$$

- 4) The prediction results of the three base learners are merged to obtain a new training set P_{New}^{Train} . The matrix P_{New}^{Train} and the original dependent variable $P_{Original}^{Train}$ are used as training data for the meta-learner.

$$P_{New}^{Train} = \begin{bmatrix} P_{Xgboost}^{1-fold} & P_{LightGBM}^{1-fold} & P_{CatBoost}^{1-fold} \\ P_{Xgboost}^{2-fold} & P_{LightGBM}^{2-fold} & P_{CatBoost}^{2-fold} \\ P_{Xgboost}^{3-fold} & P_{LightGBM}^{3-fold} & P_{CatBoost}^{3-fold} \\ P_{Xgboost}^{4-fold} & P_{LightGBM}^{4-fold} & P_{CatBoost}^{4-fold} \\ P_{Xgboost}^{5-fold} & P_{LightGBM}^{5-fold} & P_{CatBoost}^{5-fold} \end{bmatrix} \quad (19)$$

- 5) The dimension of matrix P_{New}^{Train} is transformed into three dimensions to satisfy the QR-LSTM model input

requirements, and the new matrix $P_{New_1}^{Train}$ is obtained. The prediction results $P_{New_1}^{Train}$ of the base learners and the training set $P_{Original}^{Train}$ of the dependent variable are fed to the meta-learner QR-LSTM for training. It is worth noting that the meta-learner also uses the TPE algorithm for parameter optimization to exploit its optimal predictive performance.

2.5.2 Testing phase

The detailed flow of the testing phase is shown in Figure 6. The pseudo-code for the testing phase of the proposed model is presented in Figure 7. In the testing phase, the test dataset is fed into the stacking model to implement PV power interval prediction. It is worth mentioning that the cross-validation strategy is no longer used in the testing phase and an average strategy is introduced to deal with the multiple predictions of each base learner. The details are described as follows.

- 1) The test set is fed to each base learner for prediction. A 5-fold cross-validation strategy is used in the training phase, hence for each base learner five different models are generated after training. Therefore, each base learner is capable of obtaining five predictions. The prediction results of each model are as follows.

$$P_{Xgboost}^{Test} = [P_{Xgboost}^{Test_1}, P_{Xgboost}^{Test_2}, P_{Xgboost}^{Test_3}, P_{Xgboost}^{Test_4}, P_{Xgboost}^{Test_5}] \quad (20)$$

$$P_{LightGBM}^{Test} = [P_{LightGBM}^{Test_1}, P_{LightGBM}^{Test_2}, P_{LightGBM}^{Test_3}, P_{LightGBM}^{Test_4}, P_{LightGBM}^{Test_5}] \quad (21)$$

$$P_{CatBoost}^{Test} = [P_{CatBoost}^{Test_1}, P_{CatBoost}^{Test_2}, P_{CatBoost}^{Test_3}, P_{CatBoost}^{Test_4}, P_{CatBoost}^{Test_5}] \quad (22)$$

- 2) The 5 predictions of each base learner are averaged and 3 new matrices are obtained:

$$P_{Xgboost}^{Test-New} = \frac{1}{5} \sum_{i=1}^5 P_{Xgboost}^{Test_i}, \quad P_{LightGBM}^{Test-New} = \frac{1}{5} \sum_{i=1}^5 P_{LightGBM}^{Test_i},$$

$$P_{CatBoost}^{Test-New} = \frac{1}{5} \sum_{i=1}^5 P_{CatBoost}^{Test_i}$$

- 3) The 3 matrices are combined and used as feed-in data for the meta-learner. The matrix obtained by merging the matrices is: P_{Merge}^{Test} .

$$P_{Merge}^{Test} = [P_{Xgboost}^{Test-New}, P_{LightGBM}^{Test-New}, P_{CatBoost}^{Test-New}] \quad (23)$$

- 4) The matrix P_{Merge}^{Test} is dimensionally transformed and fed into a QR-LSTM model to achieve PV power interval prediction. PV power prediction results under different quartiles are obtained. The predicted result is $W_i = [\hat{Q}_{P_i}(\tau_1|x_i), \hat{Q}_{P_i}(\tau_2|x_i), \dots, \hat{Q}_{P_i}(\tau_L|x_i)]$.

3 Evaluation indicators for interval prediction results

Prediction Interval Coverage Probability (PICP) is an important statistic for assessing prediction interval reliability, and a larger value implies that the model predicts a more trustworthy interval.

Prediction interval normalized average width (PINAW) is an essential statistic for assessing prediction interval accuracy, and a lower value suggests that the model predicts a more accurate interval. There is a relationship between PICP and PINAW. In general, the higher the PICP, the lower the PINAW, indicating that the model is more confident in the prediction interval. Simultaneously, there is a contradictory link between PICP and PINAW. When the prediction interval is large, it is easy to attain high interval coverage probability. However, prediction intervals that are too wide cannot provide accurate uncertainty information.

PINAW, PICP and WC indicators are calculated based on Eq. 24 (25) (26), respectively:

$$PICP = (1/N) \cdot \sum_{n=1}^N S_n \quad (24)$$

where N denotes the number of data. S_n represents a Boolean function. The value of this Boolean function takes 1 when the prediction interval of the model contains the true value, otherwise, it is 0.

$$PINAW = [1/(N \cdot E)] \cdot \sum_{n=1}^N (P_{up} - P_{down}) \quad (25)$$

where N is the total number of data. E denotes the difference between the maximum and minimum values of PV power. P_{up} and P_{down} respectively represent the upper and lower bound of the interval prediction.

There is a conflicting relationship between PICP and PINAW. Therefore, by combining these two indicators, a comprehensive evaluation index is proposed. The comprehensive evaluation index (WC) is calculated using Eq. 26. The smaller the WC value, the more superior the interval obtained.

$$WC = PINAW/PICP \quad (26)$$

4 Case studies

4.1 Data sets

The data used in this study are from a photovoltaic power plant in Hebei, China. The dataset is sampled at 15-min intervals. The historical data set includes active power (P), global irradiance (GI), direct radiation (DI), temperature (T), humidity (H), wind speed (WS), wind direction (WD), and pressure (P). Most of the nighttime zero-value data were removed in this study.

4.2 Selection of model input features

It is well known that PV power output is very closely related to several meteorological factors. In order to improve the accuracy of PV power prediction, it is usually necessary to filter several meteorological variables to get the meteorological variables that show high correlation with PV power output. In this study, the Kendall correlation coefficient was used for variable correlation analysis.

The Kendall correlation coefficient is calculated as follows:

$$Ken = \frac{P - Q}{\frac{1}{2} * N * (N - 1)} \quad (27)$$

where P and Q denote the number of harmonious and discordant quantities, respectively. The denominator of the formula indicates the total number of pairs of observations.

For comparison and analysis, a heat map was drawn based on the calculated Kendall correlation coefficients, as shown in Figure 8. The numbers in this figure characterise the degree of correlation between the variables. Numbers closer to 1 indicate a higher degree of correlation, and numbers closer to 0 indicate a lower degree of correlation. A negative number indicates a negative correlation.

Figure 8 shows that global irradiance (GI), direct radiation (DI), wind direction (WD), temperature (T) and humidity(H) are the main meteorological variables affecting PV output, so these five variables are chosen as input variables for the model.

4.3 Model parameter setting and data set division

The hyperparameter search range settings for the base learners and meta-learner are shown in Table 2. The LSTM model consists of one layer of LSTM network layer, one layer of Dropout layer, and one layer of Dense layer. The number of neurons of the LSTM network layer, the dropout rate, and the activation function of the Dense layer are optimized. The number of neurons in the Dense layer is 199, i.e., the quantile takes a range of values from 0.005 to 1, and the step size is 0.005. The optimizer for LSTM model training is adam, and the batch_size is 48. The epochs for the training of the LSTM model are set to 150 and an early stopping strategy is used to avoid the overfitting problem. Each base learner uses a decision tree model, which runs faster, so its hyperparameter search time is set to 200 s. The number of hyperparameter searches for the meta-learner model is 100.

The ratio of training set, validation set and test set was 7:2:1. 100 days of data were used in this study. One sunny day and one rainy day in the test set were selected separately for each model performance comparison. The model output is the predicted PV power for the 199 quantile points of the future day. The prediction interval is constructed by selecting several of the quantile predictions.

4.4 Predictive performance comparison

In order to evaluate the prediction performance of the proposed QR-Stacking model, two benchmark models and the QR-Stacking model are developed in this paper for prediction performance comparison. The two benchmark models established in this paper are QR-LSTM and Quantile Regression-Gated Recurrent Unit (QR-GRU). In order to make a valid comparison, the benchmark models QR-LSTM and QR-GRU also use the TPE algorithm for parameter search. The search parameter setting ranges of the benchmark models QR-LSTM and QR-GRU are kept the same as those of the QR-Stacking model. To verify the generalization performance of the models, the prediction performance of the three models under several different weather conditions is compared and analyzed. It is worth noting that

the prediction performance of each model is compared at 95%, 90% and 85% confidence levels in this study.

4.4.1 Comparison of the prediction performance of the models during sunny days

The prediction intervals of the three models under sunny conditions are shown in Figures 9, 10, and 11. Figures 9, 10, and 11 show that the proposed QR-Stacking model has the highest interval coverage and narrow interval width.

Table 3 evaluates the prediction interval of each model at three confidence levels using several evaluation metrics. The prediction interval coverage of all three models at 95% confidence level can meet the requirements, i.e., the coverage rate is greater than 95%. However, the prediction interval coverage of the QR-LSTM and QR-GRU models at 90% and 85% confidence levels cannot meet the requirements. In terms of PINAW and WC metrics, the prediction interval of QR-Stacking model can provide narrower prediction intervals while meeting the interval coverage requirement. At the 95% confidence level, the WC indicator of the prediction interval of the QR-Stacking model is 11.02% and 3.21% lower than those of the QR-LSTM and QR-GRU models, respectively. The WC metrics of the prediction intervals of the QR-Stacking model are 6.54% and 6.93% lower than those of the QR-LSTM and QR-GRU models, respectively, at the 90% confidence level. At the 85% confidence level, the WC metrics of the prediction intervals of the QR-Stacking model are 2.92% and 10.73% lower than those of the QR-LSTM, and QR-GRU models, respectively.

In summary, the prediction interval of the QR-Stacking model is best in sunny days.

4.4.2 Comparison of prediction performance of various models during rainy days

The prediction intervals of the three models for cloudy and rainy days are shown in Figures 12, 13 and 14. These three plots show that the prediction interval coverage of the proposed model meets the requirements and the interval is narrower.

Table 4 evaluates the prediction intervals of each model at three confidence levels using multiple evaluation metrics. The prediction interval coverage of the 3 models can meet the requirements of each confidence level. In terms of PINAW and WC metrics, the prediction interval of QR-Stacking model can provide narrower prediction intervals while meeting the interval coverage requirement. At the 95% confidence level, the WC indicator of the forecast results of the QR-Stacking model is 11.06% and 23.72% lower than those of the QR-LSTM and QR-GRU models, respectively. At the 90% confidence level, the WC metrics of the prediction interval of the QR-Stacking model are 18.38% and 23.10% lower than those of the QR-LSTM and QR-GRU models, respectively. At the 85% confidence level, the WC metrics of the prediction interval of the QR-Stacking model are 21.41% and 24.11% lower than those of the QR-LSTM, QR-GRU models, respectively.

In summary, the QR-Stacking model has the best prediction interval during cloudy and rainy days.

5 Conclusion

In this research, a QR-Stacking model with hyperparameter optimization using TPE algorithm is proposed to improve the

reliability and acuity of PV power interval prediction. The conclusions are stated as follows:

- 1) Kendall correlation coefficient is used to screen several meteorological features. This method removes the redundant features of the input data and reduces the complexity of the model.
- 2) QR-Stacking model has more superior interval prediction performance than the two benchmark models QR-LSTM and QR-GRU. QR-Stacking model can reduce the width of the prediction intervals while ensuring the coverage of the prediction intervals. In other words, the prediction intervals of the proposed model are sharper while satisfying the reliability. The superior interval prediction performance of the prediction model is further ensured by using the TPE algorithm as the hyperparametric search algorithm of the proposed model.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

Author contributions

HZ and HD proposed the methodology. HZ and HD conducted the theoretical analysis as well as the simulation verification. HZ and

HD wrote the original draft, which was reviewed and edited by RJ. RJ, JL, and YL contributed to the supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (No. 51779206), Shaanxi Province Science and Technology Department (2022JM-208).

Conflict of interest

Author HZ was employed by Shaanxi Province Distributed Energy Co.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Amir, M., Deshmukh, R. G., Khalid, H. M., Said, Z., Raza, A., Muyeen, S., et al. (2023). Energy storage technologies: an integrated survey of developments, global economical/environmental effects, optimal scheduling model, and sustainable adaption policies. *J. Energy Storage* 72, 108694. doi:10.1016/j.est.2023.108694
- Dolara, A., Leva, S., and Manzolini, G. (2015). Comparison of different physical models for PV power output prediction. *Sol. Energy* 119, 83–99. doi:10.1016/j.solener.2015.06.017
- Frilingou, N., Xexakis, G., Koasidis, K., Nikas, A., Campagnolo, L., Delpiazio, E., et al. (2023). Navigating through an energy crisis: challenges and progress towards electricity decarbonisation, reliability, and affordability in Italy. *Energy Res. Soc. Sci.* 96, 102934. doi:10.1016/j.erss.2022.102934
- Gellert, A., Fiore, U., Florea, A., Chis, R., and Palmieri, F. (2022). Forecasting electricity consumption and production in Smart homes through statistical methods. *Sustain. Cities Soc.* 76, 103426. doi:10.1016/j.scs.2021.103426
- Hussain, S. A., Razi, F., Hewage, K., and Sadiq, R. (2023). The perspective of energy poverty and 1st energy crisis of green transition. *Energy* 275, 127487. doi:10.1016/j.energy.2023.127487
- Khalid, H. M., Rafique, Z., Muyeen, S., Raqeeb, A., Said, Z., Saidur, R., et al. (2023). Dust accumulation and aggregation on PV panels: an integrated survey on impacts, mathematical models, cleaning mechanisms, and possible sustainable solution. *Sol. Energy* 251, 261–285. doi:10.1016/j.solener.2023.01.010
- Liang, L., Su, T., Gao, Y., Qin, F., and Pan, M. (2023). FCDT-IWBOA-LSSVR: an innovative hybrid machine learning approach for efficient prediction of short-to-mid-term photovoltaic generation. *J. Clean. Prod.* 385, 135716. doi:10.1016/j.jclepro.2022.135716
- Ma, M., He, B., Shen, R., Wang, Y., and Wang, N. (2022). An adaptive interval power forecasting method for photovoltaic plant and its optimization. *Sustain. Energy Technol. Assessments* 52, 102360. doi:10.1016/j.seta.2022.102360
- Markovics, D., and Mayer, M. J. (2022). Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. *Renew. Sustain. Energy Rev.* 161, 112364. doi:10.1016/j.rser.2022.112364
- Mayer, M. J., and Gróf, G. (2021). Extensive comparison of physical models for photovoltaic power forecasting. *Appl. Energy* 283, 116239. doi:10.1016/j.apenergy.2020.116239
- Nguyen, H. P., Liu, J., and Zio, E. (2020). A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by Tree-structured Parzen Estimator and applied to time-series data of NPP steam generators. *Appl. Soft Comput.* 89, 106116. doi:10.1016/j.asoc.2020.106116
- Rafique, Z., Khalid, H. M., Muyeen, S., and Kamwa, I. (2022). Bibliographic review on power system oscillations damping: an era of conventional grids and renewable energy integration. *Int. J. Electr. Power and Energy Syst.* 136, 107556. doi:10.1016/j.ijepes.2021.107556
- Rafique, Z., Khalid, H. M., and Muyeen, S. M. (2020). Communication systems in distributed generation: a bibliographical review and frameworks. *IEEE Access* 8, 207226–207239. doi:10.1109/ACCESS.2020.3037196
- Rao, S. N. V. B., Yellapragada, V. P. K., Padma, K., Pradeep, D. J., Reddy, C. P., Amir, M., et al. (2022). Day-ahead load demand forecasting in urban community cluster microgrids using machine learning methods. *Energies* 15, 6124. doi:10.3390/en15176124
- Viet, D. T., Phuong, V. V., Duong, M. Q., and Tran, Q. T. (2020). Models for short-term wind power forecasting based on improved artificial neural network using particle swarm optimization and genetic algorithms. *Energies* 13 (11), 2873. doi:10.3390/en13112873
- Wang, Z., Wang, C., Cheng, L., and Li, G. (2022). An approach for day-ahead interval forecasting of photovoltaic power: a novel DCGAN and LSTM based quantile regression modeling method. *Energy Rep.* 8, 14020–14033. doi:10.1016/j.egyr.2022.10.309



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Zhenbing Zhao,
North China Electric Power University,
China
Bo Xu,
Shanghai University of Electric Power,
China

*CORRESPONDENCE

Mingyong Xin,
✉ xinmy@csg.cn

RECEIVED 08 June 2023

ACCEPTED 10 October 2023

PUBLISHED 21 November 2023

CITATION

Xu C, Xin M, Wang Y and Gao J (2023), An efficient YOLO v3-based method for the detection of transmission line defects. *Front. Energy Res.* 11:1236915. doi: 10.3389/fenrg.2023.1236915

COPYRIGHT

© 2023 Xu, Xin, Wang and Gao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An efficient YOLO v3-based method for the detection of transmission line defects

Changbao Xu, Mingyong Xin*, Yu Wang and Jipu Gao

Electric Power Research Institute of Guizhou Power Grid Co., Ltd, Guiyang, China

The UAV inspection method is gradually becoming popular in transmission line inspection, but it is inefficient only through real-time manual observation. Algorithms are available to achieve automatic image identification, but the detection speed is slow, and video image processing is not possible. In this paper, we propose a fast detection method for transmission line defects based on YOLO v3. The method first establishes a YOLO v3 target detection model and obtains the *a priori* size of the target candidate region by clustering analysis of the training sample library. The training process of the model is accelerated by adjusting the loss function to adjust the learning direction of the model. Finally, transmission line defect detection was achieved by building a transmission line defect sample library and conducting training. The test results show that compared with other deep learning models, such as Faster R-CNN and SSD, the improved model based on YOLO v3 has a huge speed advantage and the detection accuracy is not greatly affected, which can meet the demand for automatic defect recognition of transmission line inspection videos.

KEYWORDS

YOLO v3 model, deep learning, fast defect detection, video recognition, transmission line defects

1 Introduction

With the continuous expansion of the scale of the power grid, the workload of line inspection has increased; meanwhile, the traditional manual inspection method is costly and inefficient, and there are certain dangers in implementation. In recent years, the State Grid and various electric power scientific research institutions have invested a lot of manpower and material resources to carry out UAV power inspection research, including UAV flight control technology, transmission line inspection aerial photography target identification, and fault detection technology research. The use of UAV inspection has the advantages of low cost, high efficiency, and a stronger ability to adapt to complex environments, and it can quickly collect image and video information on transmission lines, which greatly reduces the difficulty and danger of inspection work. Therefore, UAV inspection has a broad application prospect in transmission line inspection (YAN et al., 2017; Cao et al., 2021). The defect recognition method based on deep learning can automatically analyze unstructured data effectively and use deep learning algorithms to quickly process the images collected during inspection to achieve automatic detection and recognition of abnormal states of transmission line equipment, which is of great significance for enhancing the intelligence of power grids (LI et al., 2017; Li et al., 2021; WEN et al., 2021). UAVs' ual recognition of the collected images. In order to further improve the automation of machine patrol, many scholars have proposed image-based methods for the identification of transmission line equipment and

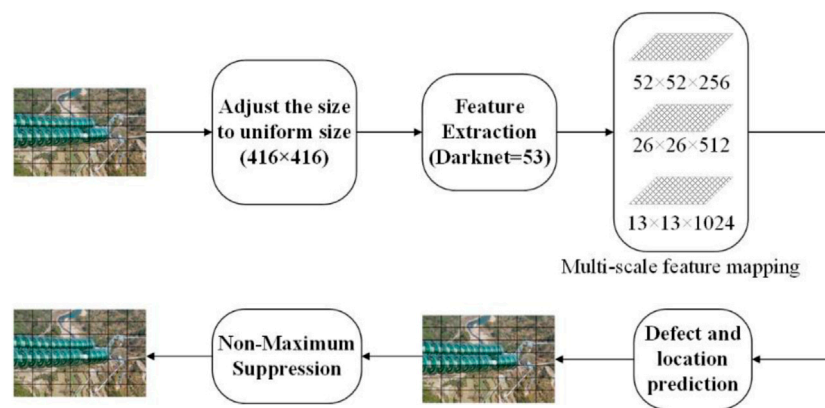


FIGURE 1
YOLO v3 model framework.

defects. Most of the algorithms currently applied for transmission line target detection need to rely on manual extraction of image features and then segmentation of targets in images by the Hough transform, Canny operator, and Gabor operator. The effectiveness of (Nguyen et al., 2018; Chen et al., 2021; Deng et al., 2021) these algorithms depend on the extraction of features, which, on one hand, is a complex task requiring strong expertise; on the other hand, there are numerous transmission line defects, which are not sufficiently expressed by manually extracted features, resulting in a single type of detection target for traditional algorithms. Some algorithms use support vector machines, neural networks, and other shallow learning algorithms to predict the target type (Yu-min et al., 2010; Cerón et al., 2014; JIANG, 2017), but they still need to extract the image features first. In recent years, machine vision technology based on deep learning has been greatly developed, and the corresponding image target detection algorithms have also achieved good performance. After the extraordinary performance of AlexNet in the image recognition competition in 2012, deep learning algorithms based on convolutional neural networks (CNNs) have become the main research direction for image classification and target detection (Li et al., 2008). Deep learning-based target detection algorithms can be divided into “dual-order method” and “single-order method”. The “dual-order method” has high accuracy, while the “single-order method” is fast, and the representative algorithms are YOLO (Simonyan and Zisserman, 2014; He et al., 2015) and SSD (Ren et al., 2017). The “double-order method” has been studied in the image detection of power system equipment (Joseph and Ali, 2016a; Redmon et al., 2016; WANG et al., 2017), but it is still in the theoretical research stage, and its detection speed is slow, which cannot meet the demand of real-time detection. In the daily UAV inspection work, a large amount of image and video data will be generated, which requires a very high speed for the target detection algorithm. Therefore, this paper establishes a defect recognition model of transmission line machine inspection images based on the YOLO algorithm using the inspection images obtained from actual engineering operation and maintenance for training (Hui et al., 2018; Lei and Sui, 2019; Wang et al., 2021). Through parameter adjustment, the practical application ability of the defect detection

model can be improved, and the real-time defect detection of the transmission line machine patrol image can be realized, which has high engineering practicability (Gong et al., 2003; Liu et al., 2020; Wang et al., 2020).

2 Inspection image defect detection model construction

The YOLO algorithm directly regresses the target location and target class at the output layer to achieve end-to-end training and detection, which is different from the original dual-order target detection method based on region recommendations. YOLO v3 adds multi-scale prediction, which makes the network more capable of detecting targets with a wide range of size variations and has higher detection speed and recognition accuracy (He et al., 2015). In this paper, we detect and identify equipment defects in transmission line inspection images based on the YOLO v3 model, and the model framework is shown in Figure 1.

For any machine patrol picture, first, the size is adjusted to a uniform size, and the picture is divided into $S \times S$ regions; then, the picture features are extracted by multiple convolution layers for each region. If the center of an electric equipment defect falls in this region, the region is responsible for predicting this defect. The center position and size of the equipment defect are adjusted by regression. The output of the model prediction is $S \times S \times (B \times 5 + C)$, i.e., $S \times S$ regions, and each region outputs B different sizes of defect prediction checkboxes and C defect type information, while for each defect prediction checkbox, there are four coordinate values and one confidence value. Finally, the model uses extreme value suppression to remove duplicate checkboxes and then predicts the actual defect types and locations of electrical equipment contained in the inspection images.

2.1 Model initialization inspection image defect detection

In this model, the input inspection images are divided into detection areas according to three scales, with 19×19 , 38×38 , 76×76



FIGURE 2
Target prediction.

× 76 detection areas. Subsequently, the nine prior defect sizes obtained from the sample library are assigned to the scale of the three detection regions, according to their size, meaning that there are three prior defect detection anchor frames in each region. For each scale detection area, if the center of the defect is within a region, that region is responsible for predicting this object, as shown in Figure 2.

2.2 Multiscale feature extraction of patrol image based on the convolutional neural network

The model directly performs feature extraction on the whole input image and achieves better detection of both large and small targets. In this study, the DarkNet-53 convolutional neural network framework is used to extract the features of inspection images, which consists of 53 convolutional layers, and each convolutional layer is followed by a linear segmentation function with leakage (Leaky ReLu) as the activation function to adapt to the nonlinear case, where five convolutional kernels have a step size of 2. The convolution result is up-sampled to obtain a multi-scale feature map. The feature extraction model is shown in Figure 3.

The model inputs a patrol image with an arbitrary RGB color pattern, and for the convolution layer, the feature vector output after the l th layer convolution operation can be expressed by Eq. 1.

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right). \quad (1)$$

Here, the range of i, j, k, l , and m depends on the structure of DarkNet-53; the range is not stated here but is intended to introduce the mathematical model of the network, where i represents the number of feature maps, j represents the number of convolution kernels, k represents the number of convolution kernels, and l represents the number of layers of the network. M_j represents the output feature map. x_i^{l-1} is the i th feature map of the $l-1$ st layer, k_{ij}^l is the j th convolution kernel, $*$ denotes the convolution operation, f (*) is the bias term, which represents the activation function, and here, the Leaky ReLu function is chosen as the convolution layer activation function, which can be expressed as follows:

$$f(x) = \begin{cases} x, & x > 0, \\ 0.1x, & x \leq 0. \end{cases} \quad (2)$$

Instead of pooling layers, this model uses convolution kernels with step size 2 in some of the convolution layers. After convolution operations with these convolution kernels, the feature vector size of the image becomes 1/4 of the original size, and the depth of the feature vector gradually deepens with the increase in the convolution window.

2.3 Defect type and location prediction based on logistic regression

This model first clusters the defect sizes of the samples before training to obtain nine priori anchor frames, and each scale feature map is responsible for detecting three scales of anchor frames. In the YOLO algorithm, the image is first cut into $N \times N$ grids, and a specified number of candidate boxes are selected for each grid, where N represents the number of meshes of the cut image. The range of N generally depends on the empirical value. The image feature vectors extracted by the DarkNet-53 convolutional neural network are input to the fully connected layer, which performs logistic regression and finally outputs a prediction vector of dimension $N \times N \times [3 \times (4 + 1 + 20)]$; it means, for three scales of feature maps, each has $N \times N$ regions, and the position information and confidence of the three

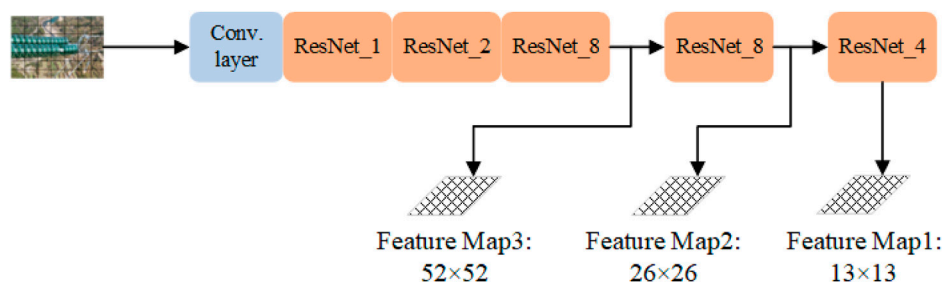
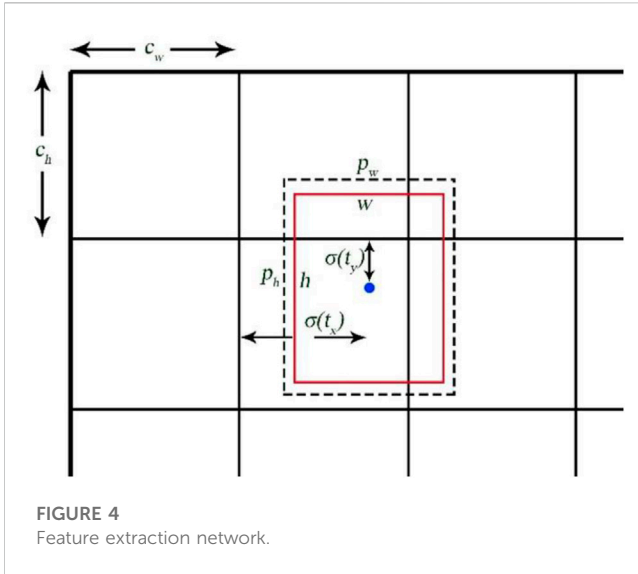


FIGURE 3
Feature extraction network.



prediction frames in each region and the defect type information form a 105-dimensional feature vector. The use of 3, 4, 1, and 20 depends mainly on the empirical values widely used after the YOLO algorithm.

2.3.1 Defect type prediction

A total of 20 equipment defect types are designed in this research, and each box uses a set of 20-dimensional vectors to represent the defect types. If the prediction is for the n th defect, the first value in the vector is 1 and the rest of the values are 0. Each prediction box has a confidence level, including the possibility of having a target in the region, the defect type, and the IOU value, as shown in Eq. 3:

$$P(\text{Class}_i|\text{Object}) * P(\text{Object}) * IOU_{pred}^{truth} = P(\text{Class}_i) * IOU_{pred}^{truth} \quad (3)$$

The model determines the possibility of a class l fault in a region based on the features extracted from the DarkNet network as $P(\text{Class}_i)$. The intersection ratio of the predicted region area to the actual region area is IOU_{pred}^{truth} . $P(\text{Class}_i|\text{Object})$ represents the probability that the target belongs to a certain defect type under the premise of the target in the box.

2.3.2 Defect location prediction

The cross-metric ratio IOU_{pred}^{truth} is used to measure the accuracy of the predicted region, which is calculated as shown in Eq. 4.

$$IOU_{pred}^{truth} = \frac{(A \cap B)}{(A \cup B)} \quad (4)$$

Here, A and B stand for the actual area and the predicted area of the equipment defect, respectively. The intersection ratio is the ratio of the intersection of the predicted area and the actual area to their concurrent set.

To overcome the instability of direct prediction, this model uses a relative position to predict the location of the defect, i.e., the predicted offset of the center of the defect relative to the top left corner vertex of the region for positioning, which is shown as in Figure 4.

The learning equation for the location of the center point is as follows:

$$\begin{cases} x = c_x + \sigma(t_x), \\ y = c_y + \sigma(t_y). \end{cases} \quad (5)$$

Here, (c_x, c_y) is the position coordinate of the upper left corner of the region, and (t_x, t_y) is the deviation of the predicted position from the actual position.

The adjustment formula of the predicted anchor box size is expressed as follows:

$$\begin{cases} w = p_w e^{t_w}, \\ h = p_h e^{t_h}. \end{cases} \quad (6)$$

Here, p_w and p_h are the size of the priori anchor frame. t_w and t_h represent the deviation of the predicted anchor frame from the actual anchor frame.

When learning the anchor box parameters, a target score is first given to the region inside the box to obtain the confidence level, and if the confidence level of the anchor box is very low, the anchor box is directly ignored. For the checkboxes with a confidence level higher than 0.5, if the mutual IOU value is high and the same object is predicted, only the anchor box with the highest confidence level is retained for learning to improve the learning speed.

2.3.3 Repeat detection target elimination based on non-extreme value suppression

With the aforementioned steps, we will get too many anchor frames, so we need to choose wisely to eliminate duplicate anchor frames. In this article, we choose the maximum suppression method to eliminate duplicate anchor frames. First, low-confidence anchor frames are suppressed, and these anchor frames most likely do not contain the target to be detected. The remaining check boxes are then categorized according to the category of the predicted defect. For targets that predict the same category of defects, first, the target with the highest confidence is selected. If the IOUs between this anchor box and other anchor boxes are higher than 0.5, anchor boxes with lower suppression thresholds and anchor boxes below 0.5 will not be processed. After that, the operation is repeated for the remaining unhidden anchor boxes. At the end of the loop, the remaining anchor frame is considered the defect target for the final prediction.

3 Model training and optimization based on the patrol inspection image sample library

This experiment is set up under the Python framework with Python version 1.5.1 using the Windows 10 operating system. Among them, CUDA version 11.4 and Python version 3.8 are used in the software environment. The hardware environment is the Windows 10 operating system, the CPU model of the testing device is 11th Gen Intel (R) Core (TM) i5-11400 @ 2.60GHz, and the GPU model is NVIDIA GeForce RTX 3060.

3.1 Training sample library construction

In this paper, 5,000 inspection images obtained from a province are used to form a training sample library, of which 4,000 images form



FIGURE 5
Example of a tower defect.



FIGURE 6
Example of fitting defects.

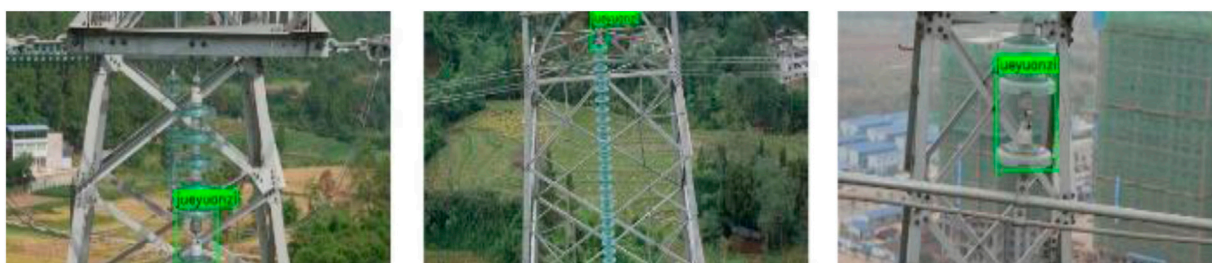


FIGURE 7
Example of insulator defects.

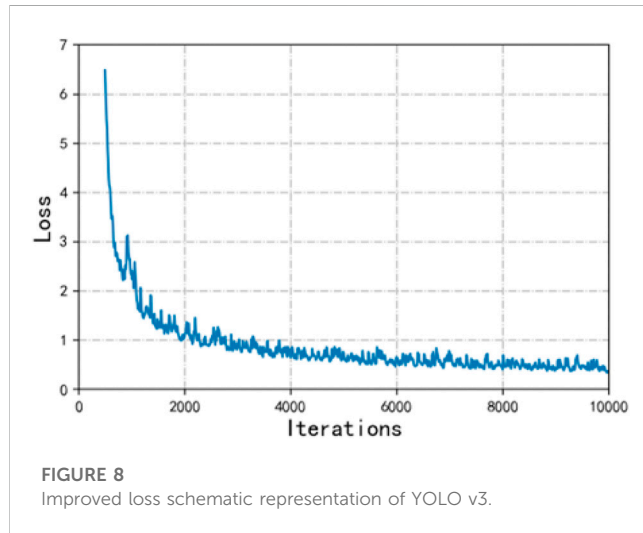
the training set and another 1,000 images form the test set. The sample images are standardized according to the Pascal VOC standard (LUO et al., 2021). The marking objects include overhead line defects, pole defects, insulator defects, and hardware defects; all the aforementioned objects constitute a sample library of typical defects of transmission lines. The size of the inspection images is not exactly the same, and considering that the shooting targets are often located in the middle of the images during the UAV inspection, in order to avoid compression and deformation of the images due to different sizes and to facilitate uniform data processing by the model, this paper crops the training images from the center to a 3:2 size and adjusts the pixels to $4,800 \times 3,200$ to remove the irrelevant parts of the edges. The defect diagrams of towers, fittings and insulators are shown in Figures 5-7 respectively.

3.2 Priori defect size selection based on a clustering algorithm

In this paper, we use a clustering algorithm to obtain the *a priori* dimensions of defects. The initial parameters of the model have an impact on the training convergence speed and training effect, and a good initial value can accelerate the convergence of the model. Through the analysis of the sample images, we found that although the fault pattern in different pictures has discrepancies, the size of the same class of faults is very close. For example, insulator faults and tower faults are generally larger in size, and the area of conductor faults is smaller. Therefore, in this study, k-clustering is used for defect sizes, and nine clustering centers are obtained. The values of these nine clustering centers are used as the sizes of the priori anchor frames of the defect detection model, which makes

TABLE 1 Priori size.

	Defect priori size
Scale 1 (19 × 19)	(512,337), (261,272), (137,144)
Scale 2 (38 × 38)	(39,281), (127,77), (63,66)
Scale 3 (76 × 76)	(47,33), (26,20), (7,9)



the selection of anchor frames better match the sizes of defects in transmission lines, speeds up the convergence of the model, and improves the accuracy of defect detection. As shown in Table 1, the prior dimensions of each anchor frame are presented.

3.3 Loss function design

The loss function of this model consists of three parts: the coordinate loss function, category loss function, and confidence loss function. The prediction results of the three scales are calculated separately and finally summed to obtain the loss function of the whole network. The coordinate loss function is expressed as follows:

$$loss_{coord} = \sum_{i=0}^{s^2} \sum_{j=0}^B l_{ij}^{obj} \left[(x_i - x_i^*)^2 + (y_i - y_i^*)^2 + \left(\sqrt{w_i} - \sqrt{w_i^*} \right)^2 + \left(\sqrt{h_i} - \sqrt{h_i^*} \right)^2 \right]. \quad (7)$$

The category loss function is written as follows:

$$loss_{class} = \sum_{i=0}^{s^2} \sum_{j=0}^B l_{ij}^{obj} (c_i - c_i^*)^2 + \lambda_{noobj} \sum_{i=0}^{s^2} \sum_{j=0}^B l_{ij}^{noobj} (c_i - c_i^*)^2. \quad (8)$$

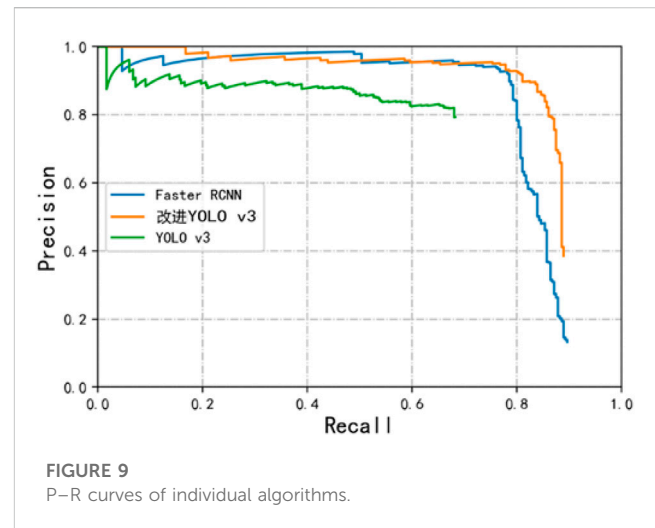
The confidence loss function is expressed as follows:

$$loss_{prob} = \sum_{i=0}^{s^2} \sum_{c \in class} (p_i(c) - p_i^*(c)). \quad (9)$$

Here, s^2 stands for the number of grid regions, B is the number of anchor frames in each region, and l_{ij}^{obj} denotes if the j th anchor

TABLE 2 Defect number statistics results.

	Training set	Test set
Hardware	2,496	611
Pole tower	1,926	516
Insulator	1,642	437
Ground wire	1,434	328
Total	7,498	1,892



frame in region i is responsible for the predicted target; if true, then it is 1, otherwise, it is 0. l_{ij}^{noobj} indicates the opposite.

3.4 Loss function weight adjustment

By analyzing the inspection images, the result shows that there are no more than four defects on most of the images, while the number of grid regions generated by each image is much larger than the number of defects. So this paper adjusts the loss weights of the regions with and without targets by parameter λ_{noobj} , and after experiments, the weight $\lambda_{noobj} = 0.1$ is chosen.

There are 20 types of defects in the design of the model, while there are only four coordinate parameters, which would result in very little influence of the coordinate parameters on the loss function if added directly. In order to increase the influence of position coordinates on the loss function to speed up the convergence, this paper adds weight $\lambda_{coord} = 5$ to the coordinate loss function. The final loss function is expressed as follows:

$$loss = \lambda_{coord} loss_{coord} + loss_{class} + loss_{prob}. \quad (10)$$

4 Model testing and result analysis

Although the YOLO algorithm has been updated to the 8th or even 9th generation versions, its essence has not changed much. Moreover,

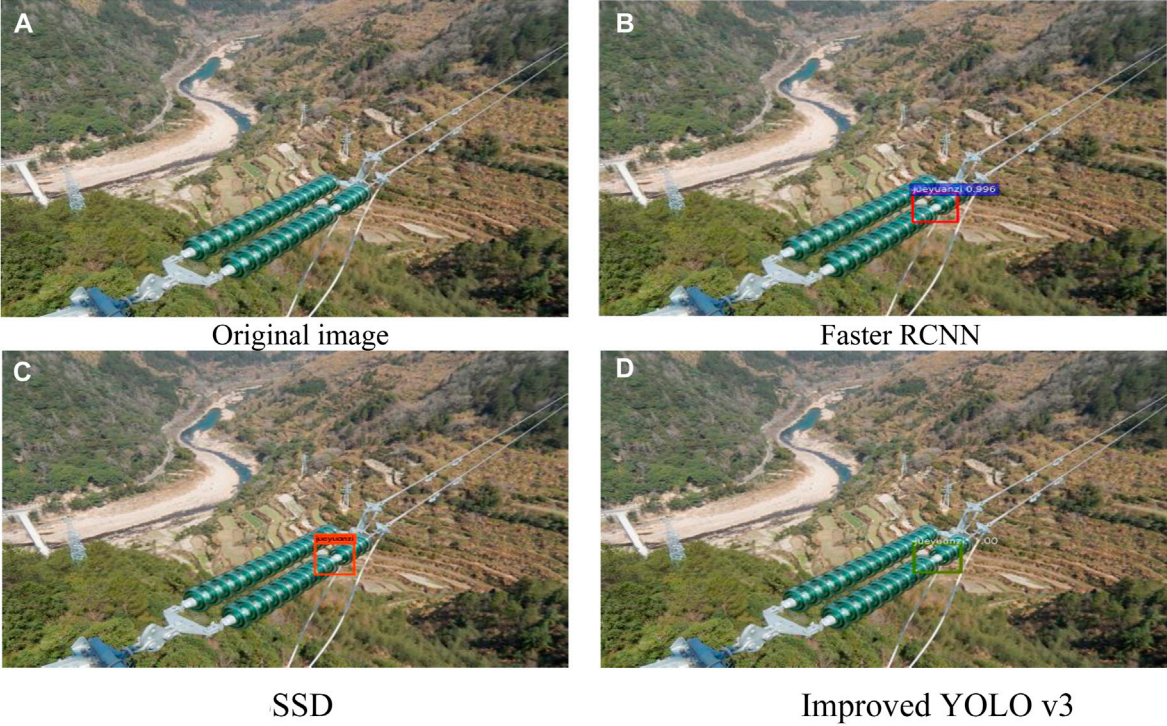


FIGURE 10
Comparison of insulator image defect detection effects.

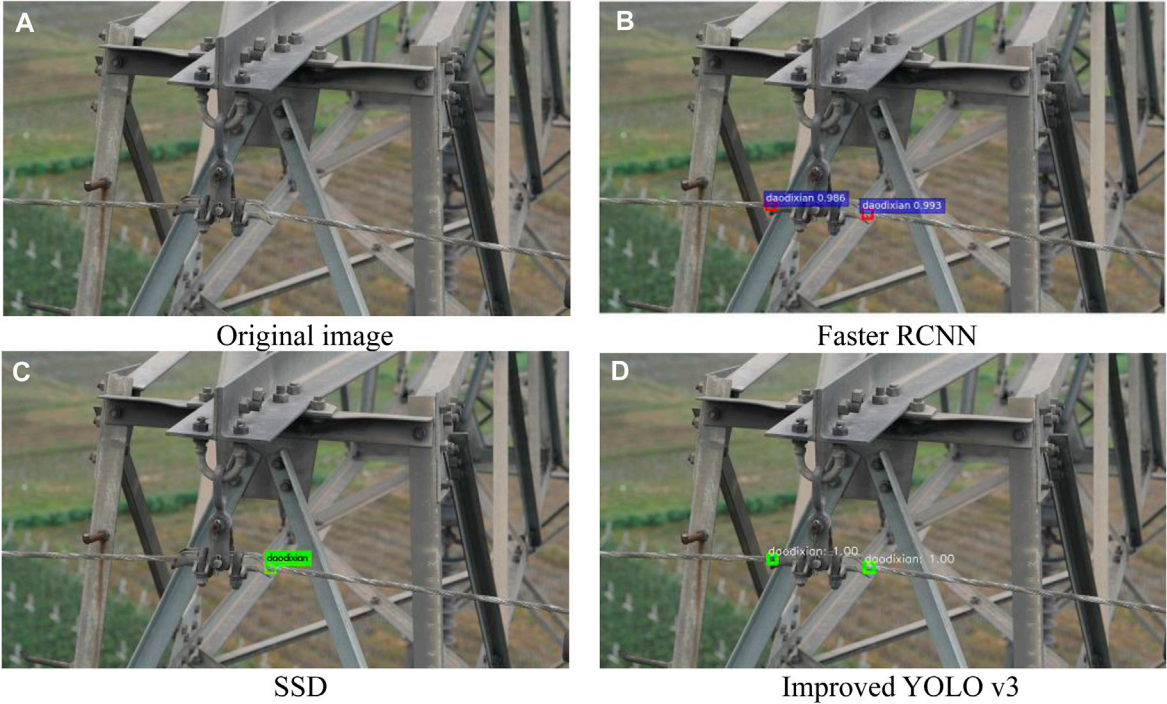


FIGURE 11
Comparison of image defect detection effects for ground wire.

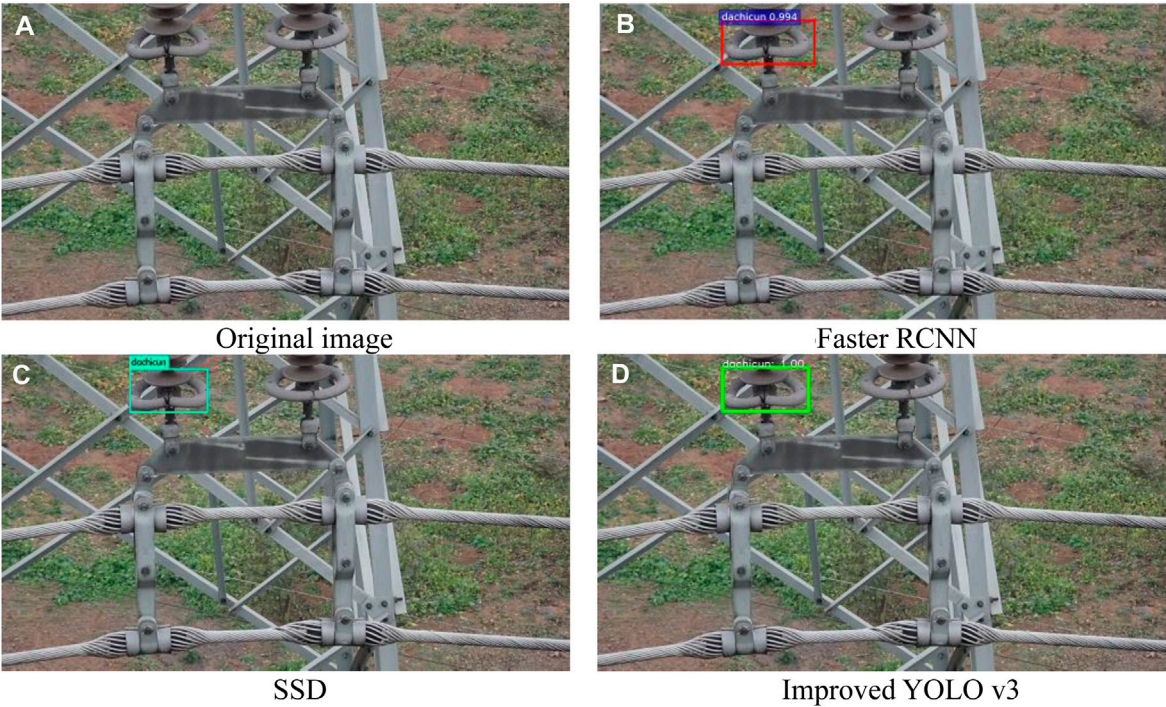


FIGURE 12
Comparison of defect detection effects on hardware image.

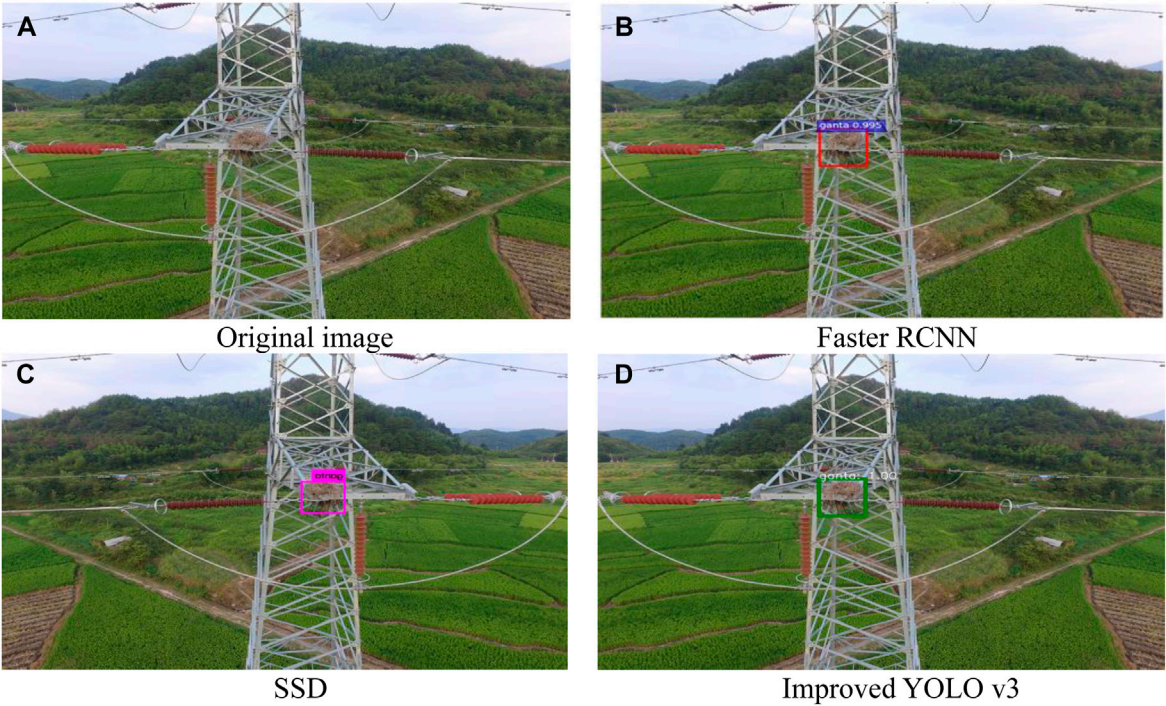


FIGURE 13
Comparison of defect detection effects on tower image.

TABLE 3 Model results' comparison.

	mAP (%)	Recall (%)	Precise (%)	Detection (ms)	GFLOPs
Faster	69.4	82.7	73.5	201	140.6
R-CNN					
SSD	66.2	76.1	72.3	102	89.5
YOLO v3	66.9	74.5	75.9	23	18.5
Improved YOLO v3	68.5	78.2	79.6	24	16.4

under the version change, the new versions all add new modules on the basis of the version of YOLO v3, thereby increasing the complexity of the model. This corresponds to making the model slower. For this reason, the most classic YOLO v3 version is used for this article. This version is different from the 1st and 2nd generation versions; not only does it have a large change in accuracy but it is also superior in speed. For this reason, the classic YOLO v3 algorithm was chosen for this paper. This section conducts simulation experiments based on the YOLO algorithm on the transmission line inspection image dataset and compares the performance with other deep learning algorithms to explore the advantages and shortcomings of the YOLO algorithm in transmission line inspection applications. Improve the change of the loss function during the training of the algorithm, as shown in Figure 8.

4.1 Defective sample analysis

The dataset used in this paper can be divided into two parts, the training and test sets. The respective types of faults contained and the corresponding numbers are shown in Table 2.

In the training set, the proportions of the four types of faults are 33.29%, 25.69%, 21.90%, and 19.12%, respectively; in the testing set, the proportions of the four types of faults are 32.29%, 27.27%, 23.10%, and 17.34%, respectively. From the statistical results, it can be seen that the number of the four types of faults is comparable, which can better meet the needs of model training and testing.

4.2 Introduction of model evaluation indicators

In this paper, the detection effectiveness of the transmission line defect detection model is evaluated by using the recall rate (recall), the precise rate (precise), and mAP (mean average precision). The recall rate is the proportion of correctly detected targets to all targets to be detected. For any detection frame with a confidence level higher than a set threshold, if the intersection ratio with a marked target is greater than 0.5 and the predicted category matches the target, the detected target is considered correct. The number of all detected correct detection frames is counted and recorded as NTP, the actual number of targets to be detected is NG, and the calculation formula for the check-all rate is shown as follows:

$$Recall = \frac{N_{TP}}{N_G} \quad (11)$$

Similarly, the number of detection frames with all confidence levels satisfying the requirement is ND, and the precise rate is calculated as follows:

$$Precise = \frac{N_{TP}}{N_D} \quad (12)$$

mAP combined with the recall rate and the precise rate is usually used as a more comprehensive indicator to evaluate a model. Improve the change of Recall during the training of the algorithm, as shown in Figure 9.

4.3 The result analysis of the patrol inspection image defect recognition

In this research, the trained model is used to conduct defect localization and identification tests on the inspection images acquired in actual operation and maintenance, and the prediction is considered accurate when the intersection ratio between the predicted target and the actual target is greater than 0.5. Several models with quality results are trained to serve as a comparison, and the experimental results are presented in Table 3.

The effects of defects detected by different algorithms are shown in Figures 10–13. Wherein Figure 10 is the insulator defect detection effect diagram, Figure 11 is the ground wire defect detection effect diagram, Figure 12 is the hardware defect detection effect diagram, and Figure 13 is the tower defect detection effect diagram. Table 2 shows that Hardwa corresponds to the label “dachicun”, PoLE Tower corresponds to “ganta”, Insulator corresponds to “jueyuanzi”, and Ground wire corresponds to “dadixian”. Based on the experimental results, it can be seen that although SSD and YOLO v3 of the single-order method are slightly inferior to Faster R-CNN of the two-order method in terms of performance, they have obvious advantages in terms of computational speed, and the prediction time of YOLO v3 is only about 1/9 of that of Faster R-CNN. Because of the two-order method, it is necessary to first show the top candidate frame and then proceed to the next step, while the single-order method directly realizes the end-to-end one-time process to complete the object detection task. So SSD and YOLO v3 have a clear advantage in speed. The double-order algorithm Faster R-CNN is slow, but its improvement on mAP is not obvious, and it is difficult to achieve fast object detection tasks. The improved YOLO v3 algorithm has only 16.4 GFLOPs. Compared to other algorithms and unimproved algorithms, the improved algorithm has a significant improvement in parallel processing speed. The performance of the improved YOLO v3 model on transmission line inspection images is also greatly improved, its accuracy far exceeded that of the Faster R-CNN model, and the recall rate is similar, but it still maintains the advantage of the single-order method in speed.

In addition to the application scenarios mentioned in this article, there are also the following scenarios:

- (1) Testing the performance of the improved YOLO v3 algorithm in different transmission line scenarios, such as different types of transmission lines and transmission lines in different environments.

- (2) Applying the improved YOLO v3 algorithm to other types of equipment, such as unmanned vehicles and helicopters, to detect the performance differences of different equipment in transmission line defect detection.

5 Conclusion

At present, the target detection algorithms applied to the defect detection of transmission lines are mainly fast R-CNN. They have high recognition accuracy but slow detection speed and are unable to realize the recognition of the large amount of video data generated in UAV inspection. This study proposes a fast detection method for transmission line defects based on YOLO v3. Its detection speed is close to 50 frames per second, which can meet the needs of video inspection. At the same time, in the dataset of this paper, the rapid detection method of transmission line defects based on YOLO v3 achieves a detection rate of 78.2% and a probability of 79.6%. Under the premise of ensuring the detection speed, it has improved compared with SSD and YOLO v3 in both indicators. The detection rate is only 5.4% lower than that of Faster R-CNN-based detection methods. At the same time, mAP is also improved compared with the single-order algorithm. Therefore, this paper argues that the transmission line detection method based on YOLO v3 can make up for the shortcomings of the Faster R-CNN algorithm and realize the rapid detection of transmission line defects based on video images.

The issues that still need to be further explored are as follows:

- (1) Insufficient diversity of the dataset: the dataset used in this article comes mainly from real-life images of transmission lines in a certain region. Although it covers a certain degree of scenarios and defect types, there may still be limitations. In order to improve the generalization ability of the algorithm, we can supplement transmission line data from other regions to increase the diversity of the dataset.
- (2) The robustness of algorithms is difficult to evaluate: transmission line defect detection often faces various complex environments and lighting conditions, so the robustness of algorithms is crucial. This article did not evaluate the performance of the improved YOLO v3 algorithm under different environments and lighting conditions.

These issues are urgent research directions, and we need to conduct further research.

References

- Cerón, A., Mondragón, B., and Prieto, F. (2014) Power line detection using a circle based search with UAV images. In *Proceeding 2014 Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, 27–30 pp. 632–639. doi:10.1109/ICUAS.2014.6842307
- Cao, Di, Hu, W., Xu, X., Wu, Q., Huang, Q., Chen, Z., et al. (2021). Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices. *J. Mod. Power Syst. Clean Energy* 9 (5), 1101–1110. doi:10.35833/mpce.2020.000557
- Chen, Y., Chen, H., Yang, J., Ma, J., and Lin, Y. (2021). Data-driven robust state estimation through off-line learning and on-line matching. *J. Mod. Power Syst. Clean Energy* 9 (4), 897–909. doi:10.35833/mpce.2020.000835
- Deng, Y., Liu, X., Jia, R., Huang, Q., Xiao, G., and Wang, P. (2021). Sag source location and type recognition via attention-based independently recurrent neural network. *J. Mod. Power Syst. Clean Energy* 9 (5), 1018–1031. doi:10.35833/mpce.2020.000528
- Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: a retrospective. *Int. J. Comput. Vis.* 111, 98–136. doi:10.1007/s11263-014-0733-5
- Gong, C., Luo, Yi, and Tu, G. (2003). COMPUTER VISION TECHNIQUE AND ITS APPLICATION TO UTOMATION OF POWER SYSTEMS. *Automation Electr. Power Syst.* 1, 76–79.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Analysis Mach. Intell.* 37 (09), 1904–1916. doi:10.1109/tpami.2015.2389824
- Hui, Li, Zhong, P., Dai, Y., and Lv, D. (2018). Study on detection method of transmission line rusty based on deep learning. *Electron. Meas. Technol.* 41 (22), 54–59.
- Jiang, Z. (2017). *Visual detection for fault diagnosis of power equipment by UAV line patrol* master's thesis. Chengdu, Sichuan, China: University of Electronic Science and Technology of China.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material; further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization: MX; methodology: MX; software: CX; validation: MX; writing—original draft preparation: JG; writing—review and editing: JG; visualization: YW; project administration: MX; funding acquisition: MX. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Intelligent Diagnosis and Common Platform for Transmission Equipment Status Based on Multi-source Visual Big Data Perception grant number (2020)2Y039.

Acknowledgments

The authors acknowledge the funding from Intelligent Diagnosis and Common Platform for Transmission Equipment Status Based on Multi-source Visual Big Data Perception (grant no. (2020)2Y039).

Conflict of interest

Authors CX, MX, YW, and JG were employed by the Electric Power Research Institute of Guizhou Power Grid Co., Ltd.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Joseph, R., and Ali, F. (2016b). Darknet: open source neural networks in C. Available at: <http://pjreddie.com/darknet/>.
- Joseph, R., and Ali, F. (2016a). YOLOv3: an incremental improvement. Available online: <https://arxiv.org/abs/1804.02767> (accessed on Apr 8, 2018).
- Lei, X., and Sui, Z. (2019). Intelligent fault detection of high voltage line based on the Faster R-CNN. *Meas. (Mahwah, N. J.)*, 138, 379–385. doi:10.1016/j.measurement.2019.01.072
- Li, J., Duan, Y., Wang, C., Wang, X., Guo, P., and Zhang, Y. (2017). Application of the unmanned aerial vehicle in the transmission line inspection. *Power Syst. Clean Energy* 33 (08), 62–65+70.
- Li, Y., Gao, W., Yan, W., Huang, S., Wang, R., Gevorgian, V., et al. (2021). Data-driven optimal control strategy for virtual synchronous generator via deep reinforcement learning approach. *J. Mod. Power Syst. Clean Energy* 9 (4), 919–929. doi:10.35833/mpce.2020.000267
- Li, Z., Liu, Y., Hayward, R., Zhang, J., and Cai, J. (2008). Knowledge-based power line detection for UAV surveillance and inspection systems. In *Proceeding 2008 23rd Int. Conf. Image Vis. Comput. N. Z.*, 26, pp. 1–6. doi:10.1109/IVCNZ.2008.4762118
- Liu, J., Jia, R., Li, W., Ma, F., Abdullah, H. M., Ma, H., et al. (2020). High precision detection algorithm based on improved RetinaNet for defect recognition of transmission lines. *Energy Rep.* 6, 2430–2440. doi:10.1016/j.egy.2020.09.002
- Liu, W. (2016). SSD: single shot MultiBox detector. *Computer vision—ECCV 2016. ECCV 9905*, 21–37.
- Luo, P., Wang, Bo, Ma, H., Ma, F., Wang, H., and Zhu, D. (2021). Defect recognition method with low false negative rate based on combined target detection framework. *High. Volt. Eng.* 47 (02), 454–464. doi:10.13336/j.1003-6520.hve.20200701
- Nguyen, V. N., Jenssen, R., and Roverso, D. (2018). Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning. *Int. J. Electr. Power & Energy Syst.* 99, 107–120. doi:10.1016/j.ijepes.2017.12.016
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: unified, real-time object detection, *Proceeding of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, NV, USA, 27–30 June 2016 Las Vegas, 779–788.
- Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Analysis Mach. Intell.* 39 (06), 1137–1149. doi:10.1109/tpami.2016.2577031
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Available at: <https://arxiv.org/abs/1409.1556> (accessed on Sep 4, 2014).
- Wang, Bo, Ma, F., Dong, X., Wang, P., Ma, H., and Wang, H. (2019). Electric power depth vision: basic concepts, key technologies and application scenarios. *Guangdong Electr. Power* 32 (09), 3–10.
- Wang, F., Ma, Ge, L., Ma, H., Wang, H., and Mohamed, M. A. (2021). Icing-EdgeNet: a pruning lightweight edge intelligent method of discriminative driving channel for ice thickness of transmission lines. *IEEE Trans. Instrum. Meas.* 70, 1–12. doi:10.1109/tim.2020.3018831
- Wang, H., Wang, B., Li, M., Luo, P., Ma, H., and Ma, F. (2020). Insulator contamination perception based on feature fusion of infrared image and meteorological parameters. *Front. Energy Res.* 9, 746378. doi:10.3389/fenrg.2021.746378
- Wang, W., Tian, B., Liu, Y., Liang, L. I. U., and Li, J. (2017). Study on the electrical devices detection in UAV images based on region based convolutional neural networks. *J. Geo-information Sci.* 19 (02), 256–263. doi:10.3724/SP.J.1047.2017.00256
- Wang, X., and Zhang, Y. (2016). Insulator identification from aerial images using Support Vector Machine with background suppression, *Proceeding of 2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, VA, USA, 7–10 June 2016 IEEE, 892–897.
- Wen, T., Zhou, D., Ming, L. I., Lin, Y., Zhao, S., Liu, J., et al. (2021). Bone mesenchymal stem cell-derived extracellular vesicles promote the repair of intervertebral disc degeneration by transferring microRNA-199a. *Comput. Eng.* 47 (03), 256–270. doi:10.1080/15384101.2020.1863682
- Wu, Xi, Yao, N., and Xu, J. (2017). Substation transformer crack image recognition based on improved neural network algorithm. *Mod. Electron. Tech.* 40 (13), 66–69.
- Yan, D., Li, Y., Guo, Q., and Yan, J. (2017). Research on transmission line UAV inspection based on image processing. In *Proceedings of the 2017 Smart Grid Development Workshop*, Beijing, China, pp. 74–77+268.
- Yu-min, G. E., Bao-shu, L. I., Shu-tao, ZHAO, and Liang, S. (2010). A method based on aerial images to detect the surface state of insulators. *High. Volt. Appar.* 46 (04), 65–68+73.



OPEN ACCESS

EDITED BY

Bo Wang,
Wuhan University, China

REVIEWED BY

Changchun Cai,
Hohai University, China
Liang Lu,
Energinet, Denmark

*CORRESPONDENCE

Wenxu Liu,
✉ 1797008102@qq.com

RECEIVED 26 August 2023

ACCEPTED 17 November 2023

PUBLISHED 29 November 2023

CITATION

Cheng X, Liu W, Bao Y and Liu X (2023),
Search strategy and line association
analysis of cascading failure accident
chain in new energy power systems.
Front. Energy Res. 11:1283436.
doi: 10.3389/fenrg.2023.1283436

COPYRIGHT

© 2023 Cheng, Liu, Bao and Liu. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Search strategy and line association analysis of cascading failure accident chain in new energy power systems

Xueting Cheng¹, Wenxu Liu^{2*}, Yueshuang Bao¹ and Xinyuan Liu¹

¹State Grid Shanxi Electric Power Research Institute, State Grid Co, Ltd, Shanxi, China, ²North China Electric Power University, Beijing, China

As the penetration rate of new energy in the power system gradually increases and the complexity of cascading faults increases, it is of great significance for the power system to comprehensively explore the chain of cascading faults in the new energy power system and quickly determine the closely related lines in the cascading faults. In response to the lack of consideration in existing research of the changes in the importance of transmission lines after the introduction of new energy, this paper proposes a cascading failure prediction index that integrates the importance and operational status of transmission lines in new energy power systems and applies it to the search for cascading failures in new energy power systems. First, the development characteristics of cascading faults were analyzed, and the main factors influencing cascading faults were identified: the importance of the transmission line and operating status of the new energy power system. Based on these factors, a prediction index for cascading faults was established, and the accident chain was searched using this index. Then, the FP-growth algorithm was used to analyze the lines in the fault chain concentration, and based on the analysis results, the correlation relationship suitable for the cascading failure lines in the new energy power system was determined. Finally, a simulation was conducted on an IEEE 10 machine 39 node system containing new energy wind turbines, and the results verified the effectiveness of the proposed indicators and strategies.

KEYWORDS

cascading failure, accident chain, new energy, line importance, association relationship

1 Introduction

In recent years, major power outages have occurred frequently worldwide, mostly caused by chain failures. As the proportion of new energy in the power system gradually increases, cascading failures have become complex. Chain failures are mainly caused by the failure of certain lines in the power grid and their withdrawal from operation, affecting the remaining branches of the power grid (Deng et al., 2022). Therefore, a comprehensive exploration of chain failures in the new energy power system and the analysis of closely related lines are of great significance for effectively preventing chain failures and major power outages.

At present, the research methods for cascading faults in power systems are mainly divided into two categories: the first type is based on complex system theory and complex network theory (Jia et al., 2016). The complex system theory evaluates the risk of cascading faults from an overall perspective by analyzing the self-organizing criticality of the power

system. The models proposed based on this theory include OPA model, CASCADE model, branch process model, and implicit fault model. These models do not focus on the physical details of the development process of cascading faults, but rather emphasize the initial conditions and macroscopic characteristics of cascading faults in the power grid, and therefore cannot describe the electrical characteristics of the power grid in detail. Actual grid operation status. Complex network theory uses metrics such as degree values, degree distributions, and betweenness centrality to describe networks and the impact of network topology on cascading faults. Models based on this theory include small world network models, Watts construction models, Holme and Kim's separated center models, Motte and Lai models, Crucitti and Latora's effective performance models, etc. These models simplify the consideration of physical processes in actual systems. Therefore, there is a gap between the actual physical process and the actual operation status of the power grid, which cannot be analyzed. The second type is based on the theory of power system analysis, and the strategy of pattern search is more closely related to the actual development process of cascading faults. For example, Li et al. proposes a fault chain model for AC/DC hybrid systems based on probabilistic power flow and short-circuit ratio theory, fully considering the impact of wind power uncertainty (Li et al., 2020). Huang et al. proposes a power grid fault assessment model considering the impact of typhoons based on existing power grid cascading fault models based on fault chains (Huang et al., 2019). Zhu et al. established a complete search model for the interlocking fault chain of AC/DC systems in large power grids, pruning the search based on the risk of line outage while ensuring accuracy and improving search efficiency (Zhu et al., 2018).

The current research on cascading faults mainly focuses on pure AC systems, considering the line distance and operating status of traditional power grids. For example, Liu et al. defines the system power flow entropy based on the entropy definition of the system and the percentage of the specified load rate components in the total number of components. The minimum load loss is obtained by taking the minimum system loss load as the objective function, and the severity index of load loss after a component failure in the i th stage of cascading faults is standardized. A system brittleness entropy index is proposed based on the combination of system power flow entropy and load loss severity index after standardization treatment. It can calculate the brittleness risk entropy corresponding to different stages of cascading faults, and be used to evaluate the impact of brittleness propagation process and component faults on the power grid (Liu et al., 2012). Qi et al. establishes a cascading failure model based on the power flow transfer factor and line topology distance and uses the entropy weight method to more comprehensively evaluate the risk of an accident chain (Qi et al., 2016). Xu et al. proposes the identification of critical power lines in the power grid based on the intermediate value of power flow (Xu and Wang, 2019). The randomness and volatility of new energy output (Wang et al., 2021) lead to uncertainty in the inline power. Zeng et al. established a line overload model accounting for fluctuations in new energy output through stochastic power flow and analyzed the risk of cascading faults in power systems containing wind power based on the overload model (Zeng et al., 2014; ATHARI and Wang, 2018). Ni et al. used the risk of line overload induced by fluctuations in new

energy as a weighting coefficient and combined it with power flow transfer entropy to propose a weak link identification method for transmission system cascading faults (Ni et al., 2019). The above research mainly focuses on analyzing the impact of the randomness of new energy output on the risk of line overload without fully considering the impact of the introduction of new energy on the importance of the line, resulting in incomplete accident chains. Therefore, it is necessary to comprehensively consider the changes in the importance and operational status of the line brought about by the integration of new energy into the power grid to obtain a more complete set of accident chains.

In response to the shortcomings of current research methods, we propose a cascading failure prediction index that integrates the importance and operational status of new energy power system lines. The main research content of this article includes the search for a chain of failures in new energy power systems and an analysis of the correlation relationships between chain failure lines. First, the main factors leading to the expansion of cascading faults were identified, and a prediction index for cascading fault accident chain routes based on the importance and operating status of new energy lines was proposed, which was used for accident chain search. Then, by obtaining the set of accident chains, the frequent pattern (FP) growth algorithm is used to analyze the lines within them, and the linear correlation relationship of cascading failures in the new energy power system is determined based on the analysis results.

2 Cascading failures and accident chain

The accident chain model originates from safety science and is made up of chains and correlations. The theory of accident chains suggests that major accidents are rarely caused by a single cause but are induced by relevant factors when multiple conditions are met simultaneously. A power system blackout accident is not caused by a single fault but rather by a chain effect of concurrent accident sequences. Assuming that the power grid has n fault chains, the set of fault chains L and the fault chains \tilde{L}_i are represented as

$$L = \{\tilde{L}_1, \tilde{L}_2, \dots, \tilde{L}_n\} \quad (1)$$

$$\tilde{L}_i = \{T_{i1}, T_{i2}, T_{im_i}\} \quad (2)$$

In the formula, T_{ij} is the j th intermediate link of the i th accident chain, where $j = 1, 2, \dots, m_i$, and the intermediate link can be a branch or a node.

The logical relationship between the system's major power outage accident, accident chain, and intermediate links in the accident chain is shown in Figure 1. The logical relationship between system power outage accidents and the set of fault chains $\{L_i\}$ is an OR gate, while the logical relationship between the fault chain and the set of intermediate links $\{T_{ij}\}$ is an AND gate. The gradual triggering of intermediate links in the same accident chain has directionality, and any triggering will lead to the triggering of the accident chain and major power outage accidents.

Accident chain triggering is a small probability event in the power grid, and the triggering probabilities of different accident chains vary greatly. The process triggered by the accident chain has different impacts on the safety of the power grid. Using risk

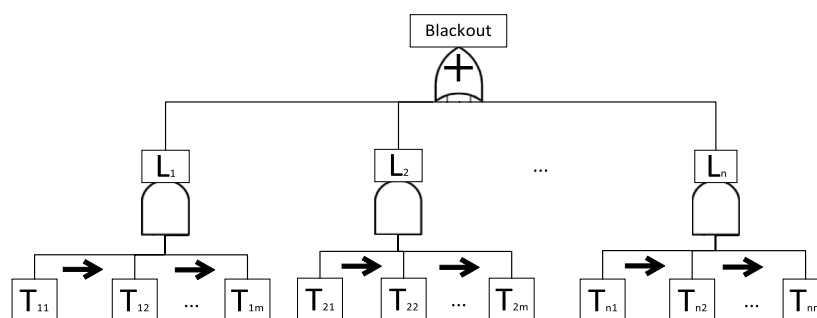


FIGURE 1
Dynamic logic diagram of a system power outage and accident chain.

assessment theory (Liu et al., 2016) to comprehensively evaluate the different characteristics between the accident chains, the risk assessment of the accident chain needs to determine the probability of occurrence of each link in the accident chain. The above set of fault chains provides all possible forms of cascading faults, but the probability of each fault chain occurring is not the same, so the criticality of lines in different fault chains also varies.

The diversity of causes and evolution modes of chain failures allows them to be classified into different types (Fang et al., 2022). The main driving factors for the evolution of chain faults include overload dominant, coordination dominant, and structural dominant. The coordinated dominant fault evolution refers to the main driving factors of major power outages, which are the unreasonable setting of secondary devices such as relay protection, low equipment reliability, etc., leading to the expansion of chain faults, caused by improper coordination of protection devices or equipment in the system; Structural dominated fault evolution refers to the significant damage to the power grid topology caused by unexpected disconnection of interconnection lines between regional interconnected power grids, premature operation of splitting devices in the early stages of chain fault development, and other factors that trigger major power outages. This is caused by structural issues or design defects in the system. The most important type is the overload-dominated type, and its evolution process is as follows: when certain components of the power system are disturbed/faulty and exit operation, the power flow passing through the faulty line will transfer to the surrounding line. If the surrounding line is affected by the overload protection action of the power flow, this will cause a new round of component removal and power flow transfer, and the above process will be repeated until a major power outage occurs. In this evolutionary mode, the frequency and voltage indicators of the power system have relatively small changes and have little impact on cascading faults. Overloading of power flow and the removal of certain lines are the main driving forces for the evolution of cascading faults.

A new type of power system with new energy as the main body, large-scale wind and solar power is replaced by conventional units through inverter grid connection, which reduces system inertia, reduces disturbance resistance, and increases the difficulty of frequency control. The asynchronous power grid interconnected by high-voltage direct current transmission between large regions has improved the security of cross regional power grids, but the

support capacity of AC power grids in each region has decreased, and frequency stability has become more prominent. The power system is always subjected to external disturbances during operation, which can lead to power imbalance, transient frequency response, and significant spatiotemporal distribution characteristics. When the electrical distance between new energy and synchronous machines is relatively close, the voltage support characteristics of the power grid are strong, and new energy and reactive power compensation overvoltage do not play a dominant role. The problem of new energy machine terminal voltage exceeding the limit caused by the power angle swing characteristics of synchronous machines is more serious. At the same time, after a power grid failure, the power grid experiences continuous DC commutation failure or locking, resulting in a change in the grid structure and a shift in power flow. After the transfer of power flow in the power grid, it may cause long-term overload or transient low voltage and high current of the AC line, causing the AC line to be cut off; It may also cause local low voltage or local high voltage, and new energy units may be disconnected due to high voltage crossing failure or low voltage crossing failure.

With the large-scale integration of new energy into the power system, cascading faults in power systems containing new energy exhibit characteristics different than those of traditional power systems. On the one hand, new energy equipment has characteristics such as strong randomness and uncertainty, and lines with new energy equipment are more susceptible to faults in other lines, which may lead to chain failures. On the other hand, the different positions of new energy units in the power grid can cause changes in the importance of the line, making searching for chain failures and accident chains more complex. In summary, this article proposes a cascading failure prediction index that integrates the importance and operational status of new energy power system lines and designs a fault chain search strategy.

3 Cascading failure accident chain prediction

3.1 Establishment of the initial fault set

The identification of the initial fault link is the primary step in establishing a cascading failure accident chain model. For the new energy power system, an initial fault indicator is established based on

the distance between the line and the new energy unit and the impact index of power flow transfer entropy. By selecting a branch that is prone to subsequent faults to be the initial fault branch, the speed and accuracy of initial fault identification can be improved.

a) Transmission capacity of line i . Each line in the power grid has different structural importance, with lines closer to new energy units having higher structural importance. Line faults with higher structural importance are more likely to lead to cascading faults. To measure the importance of transmission lines in the topology of new energy power systems, this article defines the transmission capacity of power and load nodes on (s, t) branch i as:

$$k_{i \in (s,t)} = \frac{C_i L_{\min}}{a_{i \in (s,t)} L_{\max}^{\text{line}}} \quad (3)$$

In the formula, C_i is the maximum transmission capacity including branch i ; L_{\min} is the shortest transmission distance between branch i and new energy units; and L_{\max}^{line} is the length of the line with the longest transmission distance in the power grid. When there is a change in power between the power source and the load node pair (s, t) , the power change $a_{i \in (s,t)}$ including branch i reflects the contribution of branch i to the power load node pair, which is:

$$a_{m \in (s,t)} = \frac{X_{ms} - X_{ms} - X_{mt} + X_{nt}}{x_{mn}} \quad (4)$$

The formula: X_{ms} represents the values of the m th row and s th column in the reactance matrix of the power network nodes, and $X_{ns}X_{mt}$, X_{nt} are defined similarly; x_{mn} is the reactance value of the branch L_{mn} in the power network.

b) Entropy impact index of power flow transfer based on the distance between the line and new energy units. To measure the importance of branches in the operation status of the power system, this article combines the concept of power flow transfer entropy in reference (Cao et al., 2021) to determine the impact of branch disconnection on other branches of the system. The concept of power flow transfer entropy is as follows: first, when branch i is disconnected, branch k shares the power flow increment of branch i transfer. Then, the impact rate of branch i on branch k 's power flow transfer is defined. Then, the power flow transfer entropy of the branch is defined based on the power flow distribution entropy of the node. Finally, the vulnerability index of branch consequences is defined based on the power flow transfer entropy. When the power flow impact rate of each branch is equal, the power flow transfer entropy reaches the maximum value, the possibility of each branch crossing the limit is the lowest, and the node disturbance impact that the system bears is the smallest; When the impact of the power flow is all concentrated on a single line, the minimum entropy of the power flow transfer is 0, which is most likely to cause branch out of limit faults. The relevant formulas are defined as follows:

If branch i is disconnected, the impact value η_{ki} on the transmission margin of branch k is:

$$\eta_{ki} = \frac{\Delta P_{ki}}{P_{k\max} - P_{ko}} = \frac{P_{ki} - P_{ko}}{P_{k\max} - P_{ko}} \quad (5)$$

In the formula, $P_{k\max}$ is the maximum active power that branch k can bear; ΔP_{ki} is the amount of active power transfer shared by branch k after branch i is disconnected; P_{ko} is the initial active power of branch k ; and P_{ki} is the active power borne by branch k after branch i is disconnected.

The ratio d_k of the transmission margin influence value of branch k to the sum of the transmission margin influence values of all branches is defined as:

$$d_k = \frac{\eta_{ki}}{\sum_{k \in N} \eta_{ki}} \quad (6)$$

In the equation, N represents the set of all other branches in the system except for branch i .

Therefore, considering the influence of the transmission margin on branch i , the power flow transfer entropy H_i is:

$$H_i = - \sum_{k \in N} d_k \ln d_k \quad (7)$$

Based on the power flow transfer entropy H_i of the branch transmission margin, combined with the initial power flow P_{i0} of branch i , the shortest transmission distance L_{\min} between branch i and new energy units, and the length L_{\max}^{line} of the longest transmission distance in the power grid, the impact index of power flow transfer entropy C_i based on the distance between branch i and new energy units is defined as:

$$C_i = \frac{P_{i0} L_{\max}^{\text{line}}}{H_i L_{\min}} \quad (8)$$

According to Equation 8, the greater the initial power flow borne by branch i is, the closer it is to the new energy unit, and the smaller the power flow transfer entropy is, the greater the impact of branch i 's interruption on the system, which is more likely to cause subsequent system failures.

c) Initial fault indicator. Based on the transmission capacity of the above line and the impact index of power flow transfer entropy based on the distance between the line and the new energy unit, the initial fault index of branch i is defined as:

$$E_i = C_i \times k_{i \in (s,t)} \quad (9)$$

We set an appropriate selection threshold based on the initial fault indicator, select the branches with larger indicator values to form the key branch set, and use the key branch set that is prone to causing subsequent cascading faults as the initial fault branch set.

3.2 Prediction of intermediate links

The higher the penetration rate of new energy in the power grid is, the stronger the dependence of the power grid on new energy, and the fluctuation of the output of new energy units leads to uncertainty in the transmission power of the line. Therefore, the more new energy units connected to line i , the greater the probability of line i failure. The intermediate number of lines refers to the number of times the transmission line passes through the shortest path formed between all generator buses and load buses in the power grid, which can reflect the importance of the transmission line in the topological structure of the power grid. The formula is as follows:

$$b_i = \frac{\sum_{k \neq j \in V} N_{kj}(i)(1 + \beta)}{\sum_{k \neq j \in V} N_{kj}} \quad (10)$$

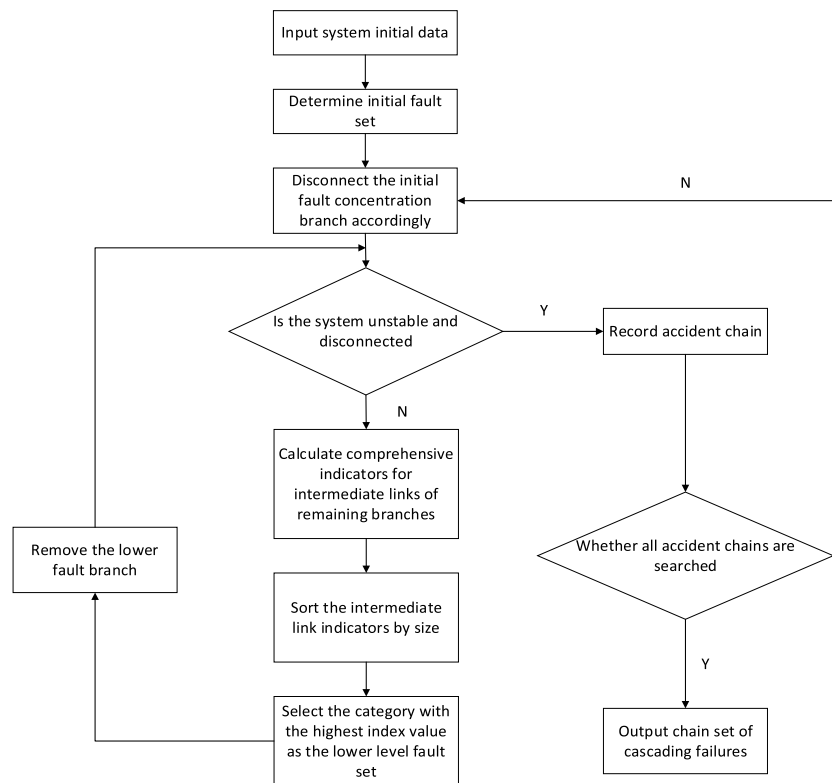


FIGURE 2
Flow chart of accident chain generation.

In the formula, i is the line number, $\sum_{k \neq j \in V} N_{kj}$ is the number of shortest paths formed between all generator buses and load buses in the network, V is the set of nodes in the network, $\sum_{k \neq j \in V} N_{kj}(i)$ is the number of times the transmission line i passes through the shortest path between all generator buses and load buses in the network, and β is the ratio of new energy power to the total power of the grid in the input power of the generator bus.

The load rate γ_i of line i is defined as:

$$\gamma_i = \left| \frac{P_i}{P_{i, \max}} \right| \quad (11)$$

In the formula, P_i is the power borne by the line.

By quantifying the above-influencing factors and combining them with the proportion of new energy equipment, the intermediate link prediction indicators can be obtained:

$$D_i = \omega_1 \gamma_i + \omega_2 b_i \quad (12)$$

In the formula, ω_1 and ω_2 are the weights of the line correlation coefficient, and the line dielectric constants are generally taken as $\omega_1 = 0.5$ and $\omega_2 = 0.5$.

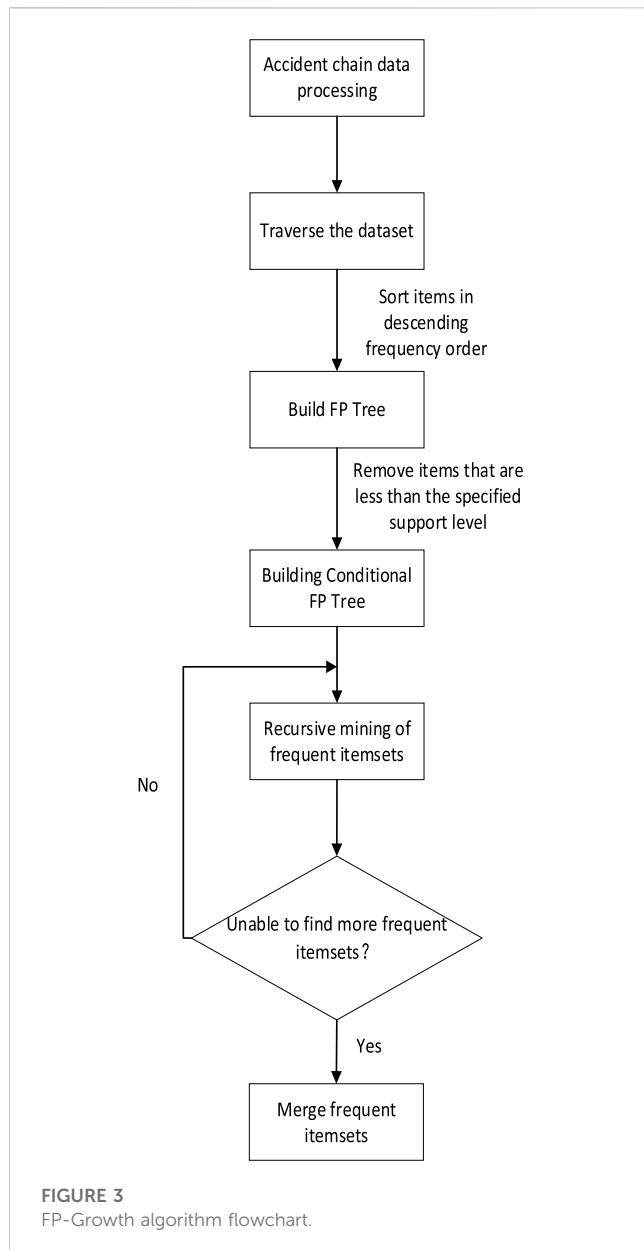
3.3 Criteria for the end of the accident chain search

When studying major power outages both domestically and internationally, we found that the process of major power outages caused by cascading faults is often accompanied by phenomena such

as power line overload and bus voltage fluctuations. These phenomena ultimately lead to instability or system disconnection in the entire power system. After the system is disconnected, the power balance between various subsystems is disrupted, leading to a series of voltage and frequency fluctuations. In severe cases, this may result in the inability of the power flow to converge or lead to large-scale power outages. In addition, to ensure the continued operation of various subsystems, it is usually necessary to cut off some generator sets and loads, but these operational measures may further expand the power outage range. Notably, even if there is no significant load loss during system disconnection, the power network will still fall into a relatively dangerous operating state, which poses a potential risk of major power outages. Therefore, system disconnection or nonconvergence of power flow has been recognized as a criterion for power outage accidents. The process of generating a chain of failures is shown in Figure 2.

4 Association analysis

We analyze the correlations between multiple lines in the accident chain set of cascading faults, and the information contained in the accident chain set is discovered. The correlation here refers to the frequent occurrence of certain lines in various development modes of chain failures, which exhibit a relatively close connection between the front and back in the accident chain and exhibit certain regularity. Many scholars at home and abroad have researched this topic and proposed many theories and algorithms



for correlation analysis. This article uses the FP-growth algorithm to explore the correlation in the accident chain.

First, we explain the relevant concepts in correlation analysis:

a) Item or candidate: An item in the database that has a unique identifier. Assuming the database is composed of m attributes (A, B, \dots, M) , then the first order itemset is $\{\{A\}, \{B\}, \dots, \{M\}\}$. The second order term set is formed by pairing elements of the first order term set, namely, $\{\{A, B\}, \{A, C\}, \dots, \{A, M\}, \{B, M\}, \dots, \{L, M\}\}$, and so on.

b) Support: Refers to the percentage of the number of supports XY corresponding to the antecedents and antecedents of a rule and the total number of AllSamples recorded.

$$\text{Support}(X, Y) = P(XY) = \frac{\text{number}(XY)}{\text{num}(\text{AllSamples})} \quad (13)$$

c) Frequent term: Refers to k -order candidate options that occur no less than the preset minimum support threshold in the dataset. At the same time, k -order candidates with occurrences

less than the set minimum support threshold are called nonfrequent terms.

Then, the following are the detailed steps of the FP growth algorithm, the FP-Growth algorithm flowchart is shown in Figure 3:

a) Build FP tree: First, the FP-growth algorithm traverses the dataset once, calculates the frequency of each item, and sorts the items in descending order of frequency. Subsequently, these sorted items are used to construct an FP tree. The FP tree is a compact data structure where each node represents an item, and the number of occurrences of that item in the dataset is recorded on the node. Each item and its corresponding frequency form a header table.

b) Build the conditional pattern base: For each item, the FP-growth algorithm constructs its corresponding conditional pattern base. The conditional pattern base refers to the collection of all prefix paths ending with the current item. Each prefix path corresponds to a frequency, while the other items in the path form a new itemset.

c) Recursive construction of the FP tree: For each item, a new FP tree is recursively constructed by utilizing its conditional pattern basis. This process iterates until no more conditional pattern bases can be constructed.

d) Mining frequent patterns from FP trees: By traversing the FP tree, all frequent patterns can be discovered. Starting from the root node of the tree, we gradually construct frequent patterns along different paths. Each path corresponds to a frequent pattern, and we can form a complete frequent pattern by adding each item on the path one by one.

The main advantage of the FP-growth algorithm is that it avoids the generation process of candidate sets, thereby reducing computational and storage costs and making it more efficient in processing large-scale datasets. In addition, this algorithm can fully utilize the structure of the FP tree to quickly discover frequent patterns. At the same time, it also supports recursive construction and mining of conditional pattern bases for frequent patterns.

Finally, the key path mining process is carried out as shown in Figure 4. First, we identify the initial fault and generate a set of chain failures based on corresponding indicators. Then, the data of the accident chain are formatted, and the appropriate minimum support is selected. A frequent item mining program based on the FP-growth algorithm is used to mine the frequent items of the accident chain, obtaining a set of frequent items that have a strong correlation with the initial fault. The lines contained in the frequent item set have strong correlations with the evolution of cascading failures.

5 Example analysis

On the basis of the IEEE39 node system structure, the synchronous units connected to Bus 30, Bus 31, Bus 33, and Bus 38 were replaced with doubly fed asynchronous wind turbines of the same generation capacity, and corresponding reactive power compensation and protection devices were configured for the wind turbines to obtain the IEEE39 node system containing new energy equipment, as shown in Figure 5.

The power generation and load capacity of the IEEE39 node system with renewable energy equipment are shown in Table 1. Among them, the power generation of the wind turbine connected to bus 30 is 250MW, the power generation of the wind turbine connected to bus 31 is 520MW, the power generation of the wind

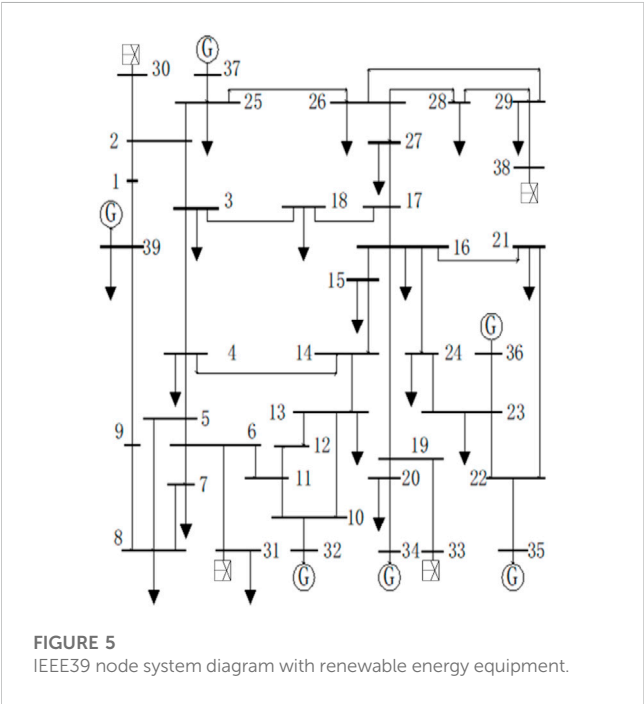
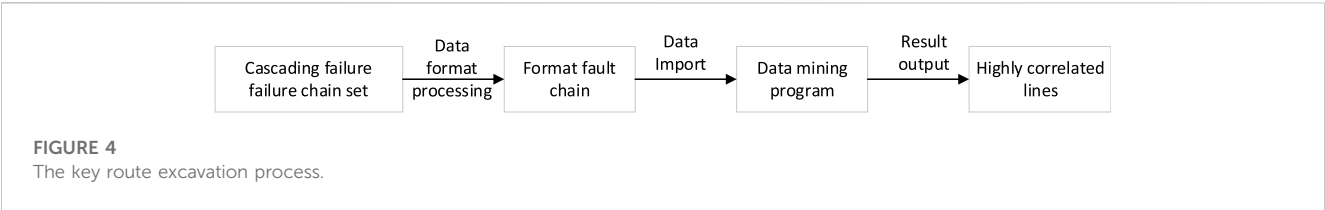


TABLE 1 Power generation and load capacity of IEEE 39 node system including renewable energy equipment.

Index	Unit (MW)
Total power generation	6,140
Coal power generation	1,068
Nuclear power generation	1,840
Other energy generation	1,000
Wind turbine power generation	2,232
Total load capacity	6,097.1

turbine connected to bus 33 is 632MW, and the power generation of the wind turbine connected to bus 38 is 830 MW. The total power generation of wind turbines is 2232 MW.

Based on the relevant formulas in section 3.1, using Python software and DlgSILENT software, the values of the transmission capacity index and the power flow transfer entropy impact index of the line are first obtained. The values of the above two indicators are multiplied to obtain the initial fault index value. The initial fault index value is listed in descending order to obtain the initial fault index of some lines, as shown in Table 2. The initial fault index threshold is set to 0.3.

From Table 3, it can be seen that after the disconnection of lines 26–28, the intermediate link prediction index of lines 21–22 is relatively large, which can be used as the next level of disconnection line. Continue to calculate the intermediate link prediction index, and obtain the relevant index values as shown in Table 4.

According to the indicators mentioned in reference (Li and Jin, 2018) and section 3 of this article, the lines of the IEEE39 node system containing renewable energy equipment were traversed. Reference (Li et al., 2020) first starts with DC power flow and combines matrix theory to propose a method for determining the power flow transfer area and calculating the power flow transfer amount based on network topology structure; Then, based on the above methods, the distribution entropy of power flow transfer and the sensitivity entropy of load impact were defined, and a comprehensive evaluation index for key lines was proposed by combining the two entropy indicators mentioned above; Finally, this indicator is used to identify key lines in power grid cascading faults. The chain of cascading faults was screened according to the same threshold, as shown in Table 5 and Table 6, respectively.

Compare the selected accident chains in Table 5 and Table 6 of the article, it can be found that under the same threshold setting, the accident chain search model proposed in this article can identify more accident chains. One reason is that the model in reference (Li et al., 2020) sets that the system will undergo splitting after some lines are disconnected, and the power flow transfer situation after the above line disconnection is no longer considered. The selection range of the initial line is small, while the model in this paper only sets a threshold for the selection of the initial faulty line, resulting in a larger search range. The second reason is that reference (Li et al., 2020) mainly considers the impact of network topology changes on power flow transfer, only proposing relevant indicators for power flow transfer, without considering the structural importance of the line itself in the power grid, and without considering the impact of new energy access on the development of cascading faults. When searching for fault chains in power grids containing new energy, it is easy to miss some lines and the obtained fault chains are not comprehensive enough.

Taking the minimum support = 3, we construct an FP tree according to the steps described in Section 4, and grow from the empty set to obtain the FP tree shown in Figure 6. The header pointer table shown in Figure 5 contains the element items contained in the dataset and their occurrence times and connects all similar element items in the FP tree through a curve. A straight line connecting nodes represents the relationship between a parent and a child, with the parent on top.

Correlation analysis is performed on the lines in the IEEE39 node cascading fault chain set containing new energy equipment using the FP-growth algorithm, as shown in Table 7.

TABLE 2 Sorting table of initial fault indicators of the line.

Line	Line transmission capacity indicators	Entropy impact index of tidal current transfer	Initial fault indicator
L_{26-28}	0.9263	0.9133	0.8460
L_{16-21}	0.7698	0.9867	0.7596
L_{01-39}	0.9685	0.4169	0.4038
L_{16-24}	0.4233	0.8326	0.3524
L_{03-18}	0.4695	0.6587	0.3093

From [Table 2](#), it can be seen that the initial fault index of lines 26–28 is much greater than 0.3, and it can be considered that lines 26–28 are prone to subsequent chain faults. So, based on the intermediate link indicators, we continue to calculate the relevant indicator values for lines 26–28, as shown in [Table 3](#).

TABLE 3 Index values of other lines after removing the initial faulty line 26–28.

Line	Line dielectric index	Line load rate	Intermediate link prediction indicators
L_{21-22}	0.312	1.023	0.1596
L_{16-19}	0.269	0.864	0.1162

TABLE 4 Index values of other lines after cutting off the line 26–28 and the line 21–22.

Line	Line dielectric index	Line load rate	Intermediate link prediction indicators
L_{22-23}	0.269	1.290	0.1735
L_{16-24}	0.302	1.139	0.1720

From [Table 4](#), it can be seen that after the disconnection of lines 26–28 and 21–22, the indicator values of lines 22–23 are relatively high. Therefore, lines 22–23 are selected as the next level of disconnection line, and the fault chain search continues until the total number of disconnected lines reaches the specified number or the power grid flow does not converge.

TABLE 5 Accident chains screened based on literature (Li and Jin, 2018) indicators.

Number		Number	
1	23–24, 21–22, 02–03, 26–27, DFIG 08, 08–09, 01–39	6	05–06, 06–07, 10–13, 02–03, 21–22, 14–15, DFIG 10
2	02–25, 21–22, 22–23, DFIG 07	7	07–08, 05–06, 04–14, 08–09, 03–04, 01–02
3	28–29, 16–19, 21–22, 16–24, 26–29, 02–03	8	01–39, 21–22, 22–23
4	01–02, 21–22, DFIG 05, 22–23	9	04–14, 05–06, 06–07, 15–16, 16–19, DFIG 10, 02–25
5	08–09, 21–22, 22–23, 02–03, 17–27, DFIG 07, 16–19		

TABLE 6 The accident chain selected based on the indicators in this article.

Number		Number	
1	26–28, 21–22, 22–23, 16–19, DFIG 05, 05–06, 06–07, DFIG 07, DFIG 10, DFIG 08	7	17–27, 21–22, 23–24, DFIG 07, 16–19, DFIG 05, 02–03, 02–25, DFIG 08
2	16–21, 23–24, DFIG 07	8	17–27, 21–22, 23–24, DFIG 07, 02–03, 10–13, DFIG 10, DFIG 08, 26–28
3	01–39, 21–22, 22–23, 09–39, 16–19, DFIG 05, DFIG 07, DFIG 10, DFIG 08, 04–05	9	15–16, 21–22, 22–23, 16–19, DFIG 05, 05–06, 06–07, DFIG 07, DFIG 10, DFIG 08
4	16–24, 21–22, DFIG 07	10	16–19, DFIG 05, 02–03, 21–22, 26–27, 23–24, 01–39
5	03–18, 21–22, 22–23, 16–19, DFIG 05, 13–14, 04–05, DFIG 07, DFIG 08, DFIG 10	11	05–06, 21–22, 16–24, DFIG 07, 10–13, 10–11, 16–19, DFIG 05, DFIG 10, DFIG 08
6	07–08, 05–06, 13–14, 21–22, 23–24, DFIG 07, DFIG 05, DFIG 08, DFIG 10, 08–09	12	14–15, 21–22, 16–24, DFIG 07, 05–06, 06–07, 04–14, DFIG 08, DFIG 05, DFIG 10

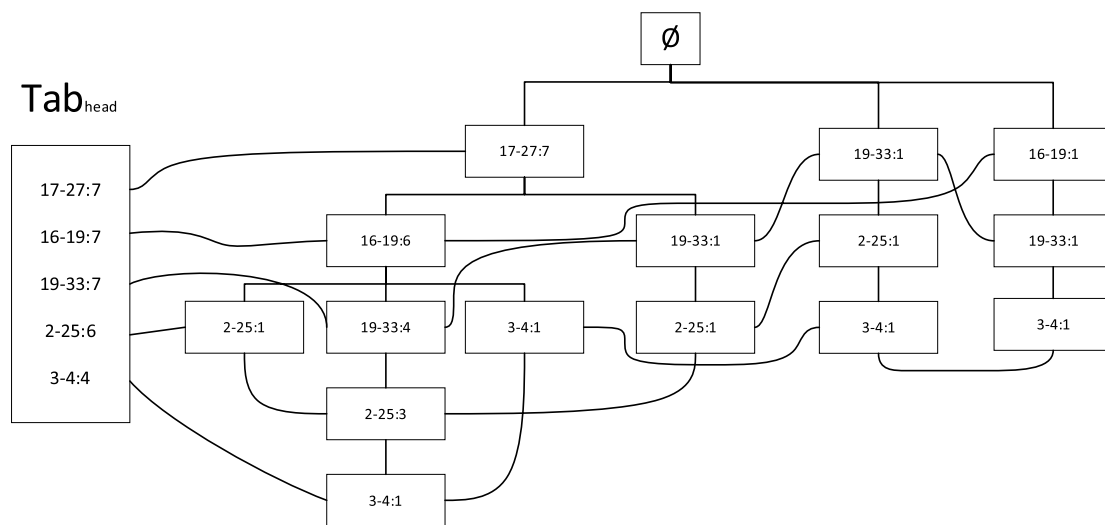


TABLE 7 Accident chain set correlation analysis results of IEEE 39 node system including new energy equipment.

1st order frequent term	2-30, 2-25, 25-37, 1-39, 17-27, 16-19, 6-31, 19-33, 22-35, 21-22, 3-4, 5-6
2nd order frequent term	2-30, 1-39; 17-27, 19-33; 2-25, 3-4; 22-35, 21-22; 17-27, 16-19; 3-4, 5-6
3rd order frequent term	17-27, 16-19, 19-33; 2-25, 3-4, 5-6

The correlation analysis results of the accident chain set are obtained by using the FP tree algorithm to search for frequent itemsets, with a minimum support of 3 set for frequent itemsets. Firstly, combined with Figure 6, search for the prefix path of line 2–25. In the figure above, search for the first 2–25 from left to right. By tracing up to the root node, you can obtain the first prefix path as {16–19, 17–27}. Then, search for the second 2–25, and then trace up to the root node to obtain the second prefix path as {19–33, 16–19, 17–27}. Then, search for the third 2–25, and then trace up to the root node to obtain the third prefix path as {19–33, 17–27}. Finally, search for the fourth 2–25, and then trace back to the root node to obtain the fourth prefix path as {19–33}.

Then, based on the support of each prefix path in the search line 2–25 mentioned above, it can be concluded that the support of prefix paths {16–19, 17–27} is 1, {19–33, 16–19, 17–27} is 3, {19–33, 17–27} is 1, and {19–33} is 1. By adding the support of each line in the prefix paths above, it can be concluded that the support of lines 16–19 is 4, lines 17–27 are 5, and lines 19–33 are 5. Due to the fact that the support of these three lines is greater than the minimum support of 3, these three lines can be used as a combination in the third order frequent term set.

By analyzing the correlated lines in the table, we find that in the IEEE39 node system, frequent items such as lines 2–30, 6–31, and 19–33 are the outgoing lines of new energy generators. Frequent items such as lines 17–27, 16–19, 21–22, and three to four are load supply lines. In the IEEE39 node system, lines 17–27, 16–19, and 19–33 are a set of transmission cross-sections. These lines have a strong correlation with the evolution of cascading faults in power

systems containing new energy, which can easily lead to large-scale power outages. Therefore, certain measures should be taken to ensure the safe and stable operation of these lines.

After calculation and simulation, 30 accident chains were obtained, and the probability table of closely related lines appearing in the same accident chain is shown in Table 8. At the same time, any combination of 7 lines was selected to obtain the probability of appearing in the same accident chain as shown in Table 9. From Table 8, it can be seen that the probability of closely related line combinations appearing in the same accident chain is higher than 50%, while in Table 9, the probability of any combination of lines appearing in the same accident chain is mostly lower than 50%, with only one combination having a slightly higher probability than 50%. This indicates that during the occurrence of chain failures, these line combinations have a higher probability of consecutive failures. By comparing the calculated line combinations with the line combinations obtained using the FP Growth algorithm, it can be found that the combination of tight lines is roughly the same, verifying the correctness and effectiveness of the FP Growth algorithm.

Calculate the probability of each line appearing in each accident chain searched in this article, and rank it from high to low to obtain the probability of some lines appearing as shown in Table 10. The more times a line appears in the accident chain, the easier it is for chain faults to spread to the line or cause other line faults, and the higher the criticality of the line. Lines with a probability of occurrence higher than 0.65 are designated as critical lines. It can be seen that the lines in the table are all critical lines, and adding

TABLE 8 Probability of closely related line combinations appearing in the accident chain.

Closely related line combinations	The probability of appearing on the same accident chain
02–30, 01–39	0.520
17–27, 19–33	0.863
02–25, 03–04	0.905
22–35, 21–22	0.556
17–27, 16–19	0.883
03–04, 05–06	0.895
17–27, 16–19, 19–33	0.503
02–25, 03–04, 05–06	0.556

TABLE 9 Probability of occurrence of any combination of lines in the accident chain.

Any combination of lines	The probability of appearing on the same accident chain
25–26, 07–08	0.095
04–05, 28–29	0.135
22–23, 01–02	0.188
02–25, 14–15	0.266
17–18, 05–06	0.505
25–26, 07–08, 03–04	0.036
22–23, 01–02, 10–11	0.163

TABLE 10 Probability of partial lines appearing in the accident chain.

Line	Probability
22–23	0.830
16–19	0.789
17–27	0.753
16–24	0.733
19–33	0.652

certain protective measures to these lines can reduce the occurrence of large-scale chain failures.

From [Table 11](#), compared with the results of references ([Xu and Zhi, 2016](#); [Shan et al., 2018](#)), the identification results of this article have more than half of the same lines, which verifies the rationality of the fault chain search in this article. Reference ([Shan et al., 2018](#)) first defines the unit entropy comprehensive

load rate by integrating the comprehensive load rate and power flow entropy of the power grid; Then, based on the HITS algorithm, the importance index of power grid nodes was defined, and a line outage consequence evaluation model was established by combining the unit entropy comprehensive load rate and the importance index of power grid nodes; Then, the probability of line breaking due to power flow exceeding the limit was defined, and a relative probability evaluation model for line breaking was constructed based on the fault chain of the line itself; Finally, based on the consequences of line disconnection and the probability of line disconnection, the fragile line index of the power grid was defined. Reference ([Shan et al., 2018](#)) suggests that the power flow impact on the line comes from random load fluctuations, without considering the importance of the line in terms of the new energy topology structure relative to the entire network, resulting in the omission of lines 4–14, 4–5, 22–23, and 16–24. These lines can break under N-1 accident conditions. Reference ([Xu and Zhi, 2016](#)) uses the DC power flow method to quickly estimate the power flow increment and load rate of other

TABLE 11 Line identification results using different methods.

Method	Line identification results
Proposed method	17–27,4–5,22–23,16–24,16–19,19–33,2–25,3–4,5–6,6–11,10–13,23–24,10–11,4–14,4–5
Reference Shan et al., 2018 method	19–33,2–25,3–4,5–6,6–11,10–13,23–24,17–27,16–19,10–11,13–14,26–27,21–22,15–16
Reference Xu and Zhi, (2016) method	23–24,10–11,4–14,2–3,22–23,16–24,17–27,16–17,19–33,2–25,3–4,5–6,6–11,10–13

lines in the system when the line is disconnected. The weighted power flow impulse entropy is used to reflect the influence of the target line on the transfer power flow caused by other line disconnections, and the weighted power flow distribution entropy is used to reflect the impact of the target line disconnection on the load rate distribution of the system line. By combining weighted power flow impact entropy and weighted power flow distribution entropy, a comprehensive vulnerability index for transmission lines is defined to identify vulnerable lines in power grid fault propagation. Reference (Xu and Zhi, 2016) focuses on considering the impact of line disconnections on the power flow transfer process of fault propagation, thus missing lines 4–5 and 16–19. These line disconnections have a significant impact on the uniformity of the electrical structure of the power grid. The identification results of this method indicate that the importance and operational status indicators of the integrated new energy power system can be used to search for more comprehensive cascading failure lines.

6 Conclusion

We propose a cascading failure prediction index based on the importance and operating status of new energy power system lines, conduct a fault chain search and analyze the correlation relationship between cascading failure lines using the FP-growth algorithm. The prediction indicators for cascading faults include the line transmission capacity, power flow transfer entropy impact index based on the distance between the line and new energy units, structural importance and load rate of the line containing new energy, and more factors that affect cascading faults. The initial fault link and intermediate development link of accident chain prediction are separated, and in the previous part, the initial faults with a low probability of occurrence are filtered out, accelerating the search speed of the accident chain. The latter part is based on a risk-first strategy, which leads the search process toward a higher risk of cascading failures. The method of line correlation analysis can calculate the correlation between multiple lines. In the FP-Growth algorithm, support and the relationship between before and after are considered to ensure the effective screening of line sets that are prone to cascading faults and have a strong correlation. By comparing the simulation results of this article with the results of other literature, we show that the set of accident chains obtained by the proposed indicators is more complete. Line correlation analysis can be used to effectively identify strongly correlated lines in a new energy power system.

References

- Athari, M. H., and Wang, Z. F. (2018). Impacts of wind power uncertainty on grid vulnerability to cascading overload failures. *IEEE Trans. Sustain. Energy* 9 (1), 128–137. doi:10.1109/tste.2017.2718518
- Cao, S., Zhang, X., Xiang, W., and Wen, J. (2021). A power flow transfer entropy based AC fault detection method for the MTDC wind power integration system. *IEEE Trans. Ind. Electron.* 68 (11), 11614–11620. doi:10.1109/TIE.2020.3031451
- Deng, H., Chen, X., Wu, J., et al. (2022). Research on power grid prevention strategy considering voltage stability and chain fault triggering. *Smart Power* 50 (1), 1–6.
- Fang, Y., Tao, H., Zhang, L., et al. (2022). Vulnerability analysis of transmission system cascading faults based on risk theory. *Electr. Autom.* 44 (02), 38–40.
- Huang, Z., Wang, D., Xiong, Q., Lan, J., Zhu, G., and Qin, F. (2019). “Risk assessment of grid cascading failure accident chain considering typhoon impact,” in 2019 12th International Symposium on Computational Intelligence and Design, USA, 14–15 Dec. 2019 (IEEE), 255–258. doi:10.1109/ISCID.2019.10141
- Jia, Y., Liu, R., Wang, P., and Han, X. (2016). Risk assessment of cascading failures in power grid based on complex network theory. *Automation, Robotics Vis.* 1–6. doi:10.1109/ICARCV.2016.7838704

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

WL: Writing–original draft. XC: Writing–original draft, Supervision. YB: Writing–original draft, Software. XL: Investigation, Writing–original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the State Grid Shanxi Electric Power Company Science and Technology Project (520530220006). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

Authors XC, YB, and XL were employed by Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fenrg.2023.1283436/full#supplementary-material>

- Li, R., and Jin, B. (2018). Identification of key lines in power systems based on power flow transfer distribution entropy and load impact sensitivity entropy. *Electr. Power Autom. Equip.* 38 (9), 148–154. doi:10.16081/j.issn.1006-6047.2018.09.022
- Li, Z., Ma, X., Luo, Y., Lang, Y., and Wu, C. (2020). “Risk analysis of cascading outage of large-scale wind power AC/DC hybrid system based on probabilistic load flow and short circuit ratio,” in IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, China, 12–14 June 2020 (IEEE), 2677–2682. doi:10.1109/ITNEC48623.2020.9085195
- Liu, S., Shi, X., Wang, T., Zhang, Y., and Cao, Y. (2016). “Power grid risk assessment method based on risk probability engineering and its application,” in IEEE International Conference on Power System Technology, China, 12–14 September 2022 (IEEE), 1–5. doi:10.1109/POWERCON.2016.7753980
- Liu, W., Wang, J., Xie, C., et al. (2012). A brittleness risk entropy based model for identifying the brittleness source of complex power grid interlocking faults. *Proceeding CSEE* 32 (31), 142–149. doi:10.13334/j.0258-8013.pcsee.2012.31.014
- Ni, L., Wen, J., Zhang, X., et al. (2019). Identification of vulnerable lines in wind and solar power grid based on weighted breaking probability and severity of breaking consequences. *High. Volt. Technol.* 45 (12), 4013–4020. doi:10.13336/j.1003-6520.hve.20191125029
- Qi, H., Shi, L., Sun, Q., and Yao, L. (2016). RNA-binding proteins in mouse male germline stem cells: a mammalian perspective. *IEEE Power Energy Soc. General Meet.* 5, 1–5. doi:10.1186/s13619-015-0022-y
- Shan, Z., Wang, H., Lin, G., et al. (2018). Identification of fragile transmission lines in power grids considering the relative probability and consequences of disconnection. *J. Zhejiang Univ. Eng. Ed.* 52 (3), 560–568.
- Wang, A., Zhong, W., Shen, Z., Wang, X., and Yang, Y. (2021). Summary of research on cascading failures of high proportion new energy power systems. *Annu. Meet. CSEE Study Comm. HVDC Power Electron.*, 343–350. doi:10.1049/icp.2021.2586
- Xu, D., and Wang, H. (2019). Assessment of high-risk cascading faults in large-scale wind power systems based on stochastic power flow and value at risk. *Power grid Technol.* 43 (2), 400–409. doi:10.13335/j.1000-3673.pst.2018.2344
- Xu, Y., and Zhi, J. (2016). Identification of fragile transmission lines in power grid fault propagation based on weighted power flow entropy. *Mod. Power* 33 (3), 88–94. doi:10.19725/j.cnki.1007-2322.2016.03.014
- Zeng, K., Wen, J., Chen, S., et al. (2014). Identification of critical lines under complex power grid cascading faults. *Proceeding CSEE* 34 (7), 1103–1112. doi:10.13334/j.0258-8013.pcsee.2014.07.012
- Zhu, Y., Li, W., and Liu, Y. (2018). “Cascading failure chains screening model of AC-DC hybrid power system,” in International Conference on Power System Technology, China, 12–14 September 2022 (IEEE), 958–964. doi:10.1109/POWERCON.2018.8601945



OPEN ACCESS

EDITED BY

Hengrui Ma,
Qinghai University, China

REVIEWED BY

José Baptista,
University of Trás-os-Montes and Alto
Douro, Portugal
Yushuai Li,
University of Oslo, Norway

*CORRESPONDENCE

Shanshan Wang,
✉ ryh513121837@outlook.com

RECEIVED 14 October 2023

ACCEPTED 27 November 2023

PUBLISHED 29 December 2023

CITATION

Sun H, Ren Y, Wang S, Zhao B and Yin R
(2023), Smart grid power load type
forecasting: research on optimization
methods of deep learning models.
Front. Energy Res. 11:1321459.
doi: 10.3389/fenrg.2023.1321459

COPYRIGHT

© 2023 Sun, Ren, Wang, Zhao and Yin.
This is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Smart grid power load type forecasting: research on optimization methods of deep learning models

Huadong Sun¹, Yonghao Ren², Shanshan Wang^{2*}, Bing Zhao²
and Rui Yin²

¹State Key Laboratory of Power Grid Safety and Energy Conservation, Beijing, China, ²China Electric Power Research Institute Co., Ltd., Beijing, China

Introduction: In the field of power systems, power load type prediction is a crucial task. Different types of loads, such as domestic, industrial, commercial, etc., have different energy consumption patterns. Therefore, accurate prediction of load types can help the power system better plan power supply strategies to improve energy utilization and stability. However, this task faces multiple challenges, including the complex topology of the power system, the diversity of time series data, and the correlation between data. With the rapid development of deep learning methods, researchers are beginning to leverage these powerful techniques to address this challenge. This study aims to explore how to optimize deep learning models to improve the accuracy of load type prediction and provide support for efficient energy management and optimization of smart grids.

Methods: In this study, we propose a deep learning method that combines graph convolutional networks (GCN) and sequence-to-sequence (Seq2Seq) models and introduces an attention mechanism. The methodology involves multiple steps: first, we use the GCN encoder to process the topological structure information of the power system and encode node features into a graph data representation. Next, the Seq2Seq decoder takes the historical time series data as the input sequence and generates a prediction sequence of the load type. We then introduced an attention mechanism, which allows the model to dynamically adjust its attention to input data and better capture the relationship between time series data and graph data.

Results: We conducted extensive experimental validation on four different datasets, including the National Grid Electricity Load Dataset, the Canadian Electricity Load Dataset, the United States Electricity Load Dataset, and the International Electricity Load Dataset. Experimental results show that our method achieves significant improvements in load type prediction tasks. It exhibits higher accuracy and robustness compared to traditional methods and single deep learning models. Our approach demonstrates advantages in improving load type prediction accuracy, providing strong support for the future development of the power system.

Discussion: The results of our study highlight the potential of deep learning techniques, specifically the combination of GCN and Seq2Seq models with attention mechanisms, in addressing the challenges of load type prediction in power systems. By improving prediction accuracy and robustness, our approach

can contribute to more efficient energy management and the optimization of smart grids.

KEYWORDS

smart grid, deep learning, optimization of intelligent systems, electric load type prediction, multi-source data, data analysis

1 Introduction

With the continuous development of society and the continuous growth of power demand, the power system is rapidly evolving into a more intelligent, efficient and sustainable form. This is the concept of smart grid. Smart grids are not only the future of the power industry, but also the key to solving energy problems, reducing carbon emissions and achieving sustainable development [Han et al. \(2022\)](#). In smart grids, understanding and predicting changes in electrical load types is critical. Electrical load refers to the power consumption pattern in the power system, which usually includes various types of loads such as household, industrial, commercial and agricultural [Li et al. \(2022a\)](#). Each load type has different characteristics and energy consumption patterns. Therefore, accurate prediction of load types can help power systems better plan power supply strategies, improve energy efficiency, reduce costs, and promote sustainable development.

However, the power load type forecasting task faces many challenges. First, the topology of the power system is usually very complex, including various substations, lines, and transmission towers, which results in complex correlations between power load data. Secondly, the diversity of time series data also increases the difficulty of prediction [Xu et al. \(2021\)](#). Different types of loads exhibit different characteristics over different time periods, which requires models to be able to identify and capture these characteristics. In addition, accurate load type forecasting requires consideration of multiple data sources, such as power system topology, historical time series data, etc. How to effectively integrate these data is also a challenge.

To address these challenges, this study focuses on developing a comprehensive deep learning approach to improve the accuracy and robustness of electric load type forecasting. We will combine graph convolutional networks (GCN) and sequence-to-sequence (Seq2Seq) models to introduce attention mechanisms to better understand and predict different types of power loads. The core idea of this method is to effectively integrate information from different data sources so that the model can better understand the complexity and temporal changes of the power system.

Studying methods and technologies for power load type prediction is of great significance to the development of smart grids and energy management. By improving the accuracy of electricity load type predictions, it can help the power system better adapt to the diversity and complexity of energy sources. This helps achieve high reliability, efficiency and sustainability of the power system, reduces resource waste, lowers carbon emissions, and promotes the integration of renewable energy. In addition, this research also provides new technical support for the intelligence and automation of the power system, laying a solid foundation for building a more intelligent power network and social infrastructure.

In research in the fields of smart grid, power load type forecasting, and deep learning, the following models are mainly used for improvement and research and development.

Convolutional neural networks (CNN) are a model that has achieved great success in the field of computer vision, but it also plays an important role in areas such as electric load type forecasting [Bhatt et al. \(2021\)](#). The main feature of CNN is its use of convolutional layers, which enables it to automatically extract spatial features from input data without manually designing a feature extractor. This feature is particularly useful for power load data processing because power load data often contains rich timing information and volatility that differs between different load types [Li et al. \(2020\)](#). In power load type prediction, the application of CNN is mainly reflected in its excellent feature extraction capabilities. CNN can capture these local features through convolution operations to identify patterns of different load types. In addition, CNN can also build hierarchical feature representation through multi-layer convolution and pooling layers, which helps to understand the information in power load data more deeply. The wide application of CNN lies in the adjustment of its convolution kernel size and number to adapt to features of different scales and complexity. In addition, CNN can also be used in conjunction with other deep learning models and techniques, such as recurrent neural networks (RNN) and attention mechanisms, to better capture temporality and correlation between data.

Recurrent neural network (RNN) is a type of deep learning model suitable for sequence data, which is of great value in power load type forecasting tasks. The unique feature of RNN is that it has internal cyclic connections, which allows the model to process variable-length time series data, which is very important for modeling power load data. In power load type forecasting, RNN can be regarded as a sliding window in time, which can capture the dependence between load data at different time points. This is key to understanding the evolution of load types over time [Xiao and Zhou \(2020\)](#). However, traditional RNN is prone to problems such as gradient disappearance or gradient explosion on long sequence data. For this reason, improved RNN models such as gated recurrent unit (GRU) and long short-term memory network (LSTM) have emerged. GRU controls the flow and memory of information by introducing update gates and reset gates to better process time series data [Dhruv and Naskar \(2020\)](#). These improved RNN models perform well in power load type forecasting, especially when long-term dependencies need to be considered. Choosing an appropriate RNN model depends on the characteristics of the data and the requirements of the task to ensure that it can better capture the information of time series data.

Temporal convolutional network (TCN) is a model that combines CNN and RNN, and it has broad application prospects in power load type forecasting. TCN uses convolutional layers to capture the local and global relationships of time series data,

avoiding the gradient problem in traditional RNN. This makes TCN ideal for processing long sequences of data, especially when power load type forecasting needs to consider a wider range of historical information [Arumugham et al. \(2023\)](#). The main feature of TCN is that it has an extended receptive field of variable length, which means that the model can effectively capture features at different time scales. In power load type forecasting, different load types may show different patterns on different time scales, so TCN can help the model better adapt to this diversity [Fan et al. \(2023\)](#). In addition, TCN can be combined with other technologies such as attention mechanisms to further improve model performance.

Gated Recurrent Unit (GRU) is an improved RNN model designed to overcome the problems of traditional RNN. The main feature of GRU is that it has update gates and reset gates inside, which allow the model to better control the flow and memory of information [Cheon et al. \(2020\)](#). In power load type forecasting, GRU can be used to capture long-term dependencies of time series data. One of the advantages of GRU is its simplicity and efficiency. Compared with LSTM, GRU has fewer parameters and therefore trains faster [Daniels et al. \(2020\)](#). This makes GRU ideal for processing large-scale time series data. In power load type forecasting tasks, choosing the GRU model can reduce computational costs while maintaining high performance.

Deep reinforcement learning (DRL) is a powerful model whose main feature is to learn optimal decision-making strategies through interaction with the environment. In the field of smart grid, DRL can be used for load management and optimization to achieve the best balance of energy efficiency and power supply stability [Leng et al. \(2021\)](#). The DRL model can dynamically adjust the power supply strategy according to changing power load conditions, thereby improving energy utilization efficiency. Although DRL models generally require more data and computing resources, they perform well in handling complex decision-making problems. In power load type forecasting, DRL can be combined with other deep learning models to achieve higher-level decision-making and control, contributing to the development of smart grids and optimization of power systems [Huang et al. \(2019\)](#). The choice of DRL model usually depends on the complexity of the task and the problem that needs to be solved.

However, there are some shortcomings when applying these models to the study of smart grid power load type prediction problems. Although convolutional neural networks (CNN) are good at extracting spatial features, they have limited modeling of time series data and are difficult to capture dynamic changes in load types. Recurrent neural network (RNN) and its improved models (such as GRU and LSTM) can handle time series data, but are susceptible to problems such as gradient disappearance and gradient explosion, which limit their long-term dependency modeling capabilities. Although temporal convolutional network (TCN) overcomes the gradient problem of RNN, it may not be flexible enough to adapt to different scales of temporal data. Deep reinforcement learning (DRL) requires a large amount of data and computing resources, has challenges in complexity, and is not suitable for all power load type prediction scenarios.

In view of this, we propose a GCN-Seq2Seq model that integrates the attention mechanism. This model combines graph convolutional network (GCN) and sequence-to-sequence model (Seq2Seq), and introduces an attention mechanism, which has the

following advantages. First, GCN can effectively capture the complex topology of the power system and help the model understand the relationship between different load types. Secondly, the Seq2Seq model is suitable for sequence generation tasks, mapping historical time series data to load type prediction sequences, and better considering timing. Most importantly, the attention mechanism we introduced enables the model to automatically focus on the most important information, improving the accuracy of predictions. Our model has advantages in comprehensively considering the topology, time series data and correlation of the power system, and is expected to improve the performance and efficiency of power load type prediction, which is beneficial to the development of smart grids and the optimization of power systems.

The main contributions of this study are as follows:

- Proposal of new deep learning model. We propose an innovative deep learning model that combines graph convolutional networks (GCN) and sequence-to-sequence models (Seq2Seq), and introduces an attention mechanism. This model can simultaneously consider the topology and timing data of the power system and automatically capture the correlation of load types, thereby improving the accuracy and accuracy of predictions.
- Research on multi-source data fusion. We apply multi-source data fusion to the power load type prediction task, taking into account the topological information and historical time series data of the power system. This data fusion method is expected to improve the robustness and accuracy of load type forecasting and provide more comprehensive information for intelligent management of power systems.
- Promote the sustainable development of smart grids. The results of this study are expected to contribute to the sustainable development of smart grids and efficient management of power systems. Through more accurate load type forecasting, the power system can better adapt to changing demands, improve the reliability and efficiency of power supply, and also provide strong support for the development of sustainable energy integration and smart grids.

In the following sections, we summarize all the model diagrams involved in this study, as well as the data analysis diagrams in Part II. In the third part, we introduce in detail the deep learning model we proposed, that is, the GCN-Seq2Seq model incorporating the attention mechanism, and elaborate on the structure diagram and basic principles of the model. The fourth part is our experiment, which introduces the data sets used in this study, the detailed experimental settings and the analysis of experimental results. The fifth part is the conclusion and summary of the full text. We also describe the shortcomings of this study and the next research direction.

2 Related work

2.1 Intelligent power system

As an innovative field in the power industry, smart power systems cover a series of advanced technologies and concepts,

aiming to improve the intelligence, efficiency and sustainability of the power system. The basic concept includes real-time monitoring, control and optimization of power networks to better meet growing power demand. The origins of smart power systems can be traced back to the digital transformation of traditional power systems. With the continuous advancement of information technology, smart power systems have gradually evolved into a complex network that integrates elements such as advanced sensors, communication technology, data analysis, and artificial intelligence to make the power system more flexible and intelligent.

In the field of smart power and energy management, recent research demonstrates the rise of hybrid technology solutions that focus on improving operational efficiency and system resilience against potential risks. A study proposes a reinforcement learning-based energy management system designed to optimize the performance of fuel cell and battery hybrid electric vehicles [Reddy et al. \(2019\)](#). The core of the system is to dynamically adjust the distribution of electric energy, showing the possibility of improving energy efficiency under changing risk conditions. In response to smart grid security issues, especially the threat of denial of service (DoS) attacks, some research has developed a distributed control mechanism. This mechanism combines the system's communication capabilities and control responses to ensure the stability of grid dispatch and operation even in the event of a cyber attack [Li et al. \(2022b\)](#). In addition, for microgrid energy management issues, the latest research introduces a distributed energy management framework to complete dual-mode energy distribution within a predetermined time through event-triggered communication technology. This method can effectively deal with communication delays and ensure the accuracy and reliability of energy distribution [Liu et al. \(2023\)](#). These studies as a whole reflect that the methods used by intelligent systems to improve performance and security are becoming increasingly complex, and interdisciplinary technology integration is a significant trend in current development. From reinforcement learning algorithms to the application of advanced communication protocols, it reflects important steps taken in smart energy distribution and power grid management.

However, smart power systems also face some challenges. Especially in terms of power load type forecasting, challenges mainly include the complex topology of the power system, the diversity of time series data, and the correlation between data. Addressing these challenges is crucial to achieve comprehensive optimization of smart power systems and improve power load type forecast accuracy.

2.2 Deep learning technology

Deep learning technology has achieved remarkable application results in the field of power systems, providing strong support for the intelligence and efficiency of power systems. In terms of power load forecasting, deep learning algorithms can be used to learn and model historical load data to achieve accurate predictions of future power loads. In terms of power system optimization, deep learning technology is used to learn the topology structure and operating status of the power system to achieve real-time optimal dispatch of the power system [Ibrahim et al. \(2020\)](#). In terms of smart grid management, deep learning technology is

used to process a large amount of time series data in the power grid, which can realize real-time monitoring, fault detection and intelligent dispatching of the power grid. In terms of power load forecasting, deep learning technology has been successful in many cases. For example, in the power load forecasting of the State Grid, deep learning methods achieve highly accurate load forecasting by learning the complex spatiotemporal relationships of the power system, providing an important basis for reasonable dispatch of the power system [O'Dwyer et al. \(2019\)](#). In terms of power system optimization, deep learning technology has also shown strong capabilities. By training large-scale data from the power system, deep learning models can better understand the modes and trends of system operation, thereby achieving intelligent scheduling and optimization of the system.

Compared with traditional methods, deep learning technology has significant advantages. Deep learning models can learn and capture the complex spatiotemporal relationships in power systems and better adapt to the nonlinear characteristics of the system. Deep learning models can achieve end-to-end learning, learn feature representations directly from raw data, without the need to manually extract features, and improve the generalization ability of the model [Zhang et al. \(2019\)](#). The deep learning model can automatically adjust model parameters to adapt to the characteristics of different power systems, and has stronger adaptability and generalization capabilities.

Although deep learning has achieved remarkable results in power systems, it still faces some challenges. Issues such as power system complexity, data uncertainty, and model interpretability remain the focus of current research. The reason for choosing the deep learning method in this study is its advantages in processing large-scale data, learning complex relationships, and adapting to uncertainty.

2.3 Optimizing deep learning models

In terms of optimization of deep learning models, a variety of methods have emerged in recent years, especially in applications in the field of power systems, including transfer learning, reinforcement learning, hyperparameter optimization, adversarial training, etc. Transfer learning uses the knowledge learned on one task to help learn on another related task. Transfer learning can reduce the dependence on a large amount of annotated data and improve the generalization of the model [Hafeez et al. \(2020\)](#). The introduction of reinforcement learning methods allows the model to optimize its own performance through interaction with the environment, which is particularly suitable for real-time dispatch and control problems in power systems. Optimizing the hyperparameters of deep learning models through search algorithms or adaptive methods can improve the performance and robustness of the model. Introducing adversarial training enables the model to better cope with perturbations and attacks on input data, and improves the robustness of the model.

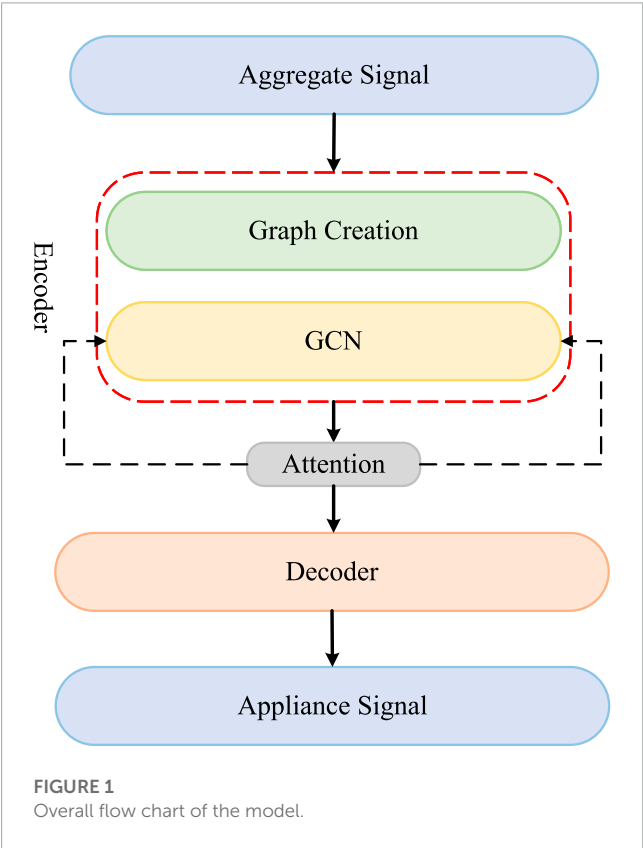
Optimization schemes based on meta-learning have been applied to deep learning models, especially in the field of power systems. This method has confirmed its effectiveness in improving model performance between different systems through the practice of transfer learning [Zhou et al. \(2020\)](#). At the same time,

reinforcement learning technology also shows great potential in load forecasting. It can enhance the model's adaptability to complex changes by reproducing different load conditions in a simulated environment. In addition, the introduction of adversarial training is regarded as an important development in the field of power system security. Adversarial samples are added to improve the system's ability to identify network attacks, thereby enhancing the defense mechanism [Ye et al. \(2020\)](#). These research results provide a wealth of ideas and methods for optimizing deep learning models, and provide a reference for our optimization of deep learning models in power load type forecasting.

3 Methodology

3.1 Overview of our network

For the power load type prediction problem, significant progress has been made in the application of deep learning technology in smart power systems and related work in model optimization. In order to further improve the prediction accuracy, this study adopts an overall model that integrates graph convolution network (GCN) and sequence-to-sequence model (Seq2Seq), and introduces an attention mechanism to solve the problem of smart grid power load type prediction. This model was chosen due to considerations of the complexity and diversity of power systems and the need for accuracy and global information capture. The basic principle of this overall model is to view the power system as a graph structure, where nodes represent specific time points of load data and edges



```
Input: Training data:  $X, Y$   
Output: Trained GCN-Seq2Seq model  
Initialize model parameters  $\Theta$  randomly;  
Initialize learning rate  $\alpha$ ;  
Initialize loss function:  $Loss = 0$ ;  
while not converged do  
  Sample a batch of training examples  $(x_i, y_i) \sim (X, Y)$ ;  
  Calculate forward pass through GCN-Seq2Seq model:  
  
       $h_i = GCN(x_i, A)$   
       $\hat{y}_i = Seq2Seq(h_i)$   
  
  Calculate loss:  
  
       $Loss_i = CrossEntropy(\hat{y}_i, y_i)$   
       $Loss = Loss + Loss_i$   
  
  Update model parameters with backpropagation:  
       $\Theta = \Theta - \alpha \nabla_{\Theta} Loss$   
  
end  
Evaluation:  
for each evaluation batch  $(x_e, y_e) \sim (X_{eval}, Y_{eval})$  do  
  Calculate prediction  $\hat{y}_e$  using the trained model;  
  Calculate Recall, Precision, and other evaluation metrics using  $\hat{y}_e$  and  $y_e$   
end  
return Trained GCN-Seq2Seq model
```

Algorithm 1. GCN-Seq2Seq Training.

represent topological relationships between nodes. First, through the GCN encoder, the model can effectively capture the topological information of the power system and represent the node features into the encoding of graph data. Next, the Seq2Seq decoder accepts historical time series data as an input sequence and generates a load type prediction sequence. In this process, an attention mechanism is introduced, allowing the model to fuse information based on the importance of different input data and better understand the relationship between time series data and graph data. The advantages of this model are obvious. First, it can comprehensively consider the topology and timing data of the power system while automatically capturing the correlation between different load types, thereby improving the accuracy of prediction. Secondly, the introduction of the attention mechanism enables the model to focus on the most important information for the current prediction, further improving the model performance. Most importantly, the comprehensiveness and global information capturing capabilities of this model are expected to provide a more powerful tool for intelligent management of power systems and forecasting of power load types.

The structure diagram of the overall model is shown in [Figure 1](#), which shows the relationship between the GCN encoder, Seq2Seq decoder and attention mechanism, forming a comprehensive power load type prediction model.

The running process of the GCN-Seq2Seq model is shown in [Algorithm 1](#).

3.2 Graph convolutional network model

In the model of this study, the graph convolutional network (GCN) is a key component used to process the topological structure information of the power system [Hossain and Rahnamay-Naeini \(2021\)](#). The basic principle of GCN is to capture the relationship between nodes in graph data through effective information transfer [Peng et al. \(2023\)](#), and then encode the features of the nodes [Chen et al. \(2022\)](#). In the overall model, the role of GCN is to treat the power system as a graph structure, in which the nodes of the graph represent load data at different time points, and the edges represent topological relationships between nodes, such

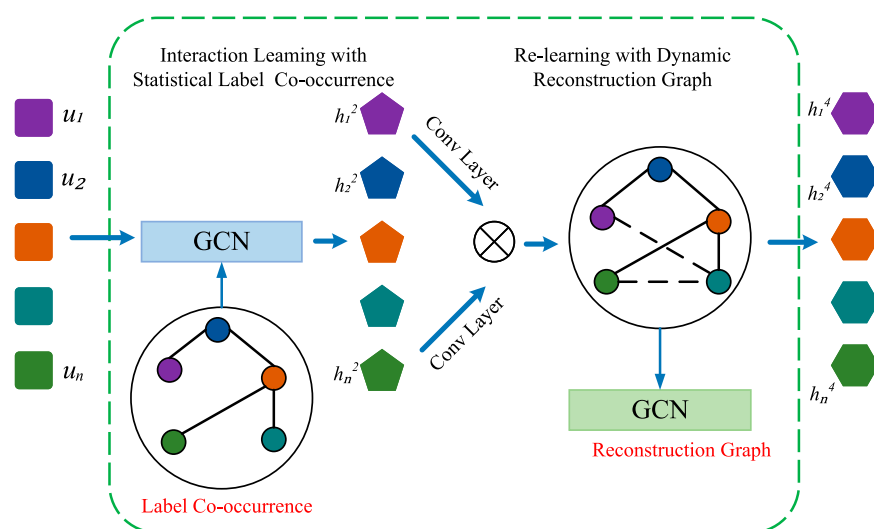


FIGURE 2
Flow chart of the GCN model.

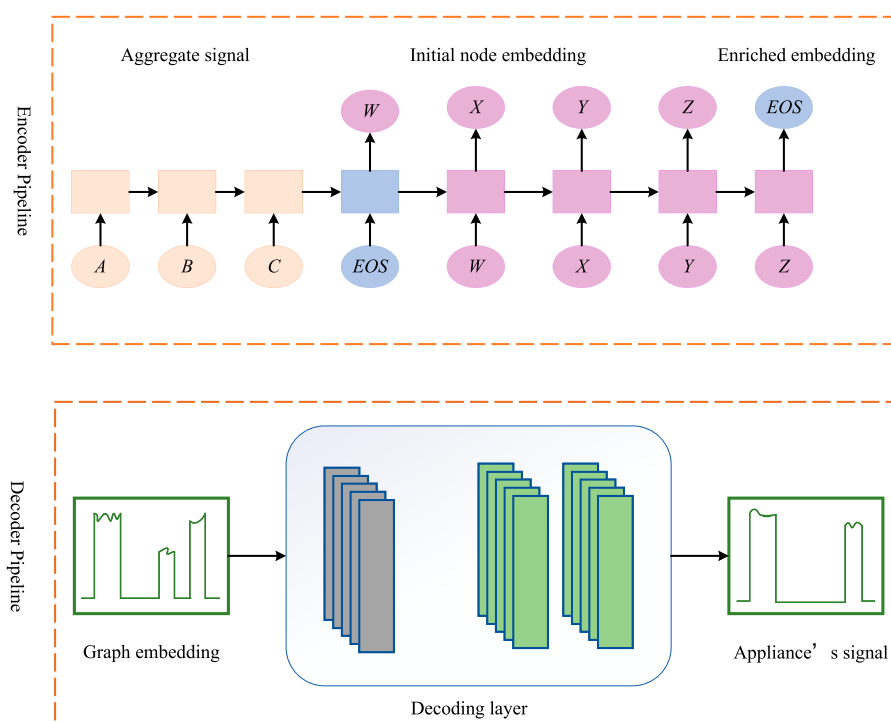


FIGURE 3
Flow chart of the Seq2Seq model.

as connection relationships. These nodes and edges constitute the topological information of the power system. The advantage of GCN in power system modeling is mainly reflected in its effective processing of complex topological structures. Compared with traditional methods, GCN can capture the relationship between nodes more comprehensively and achieve a high degree of abstraction and expression of the power system topology. Through an iterative information transfer process, GCN is able to update

the characteristics of each node to the weighted average of the characteristics of its neighboring nodes, effectively integrating topological relationships into feature representation. This enables the model to better understand the interactions and correlations between different nodes in the power system, thereby improving the accuracy of load type predictions. Specifically, the ability of GCN lies in encoding the node information of the power system so that the model can better understand the spatiotemporal relationship

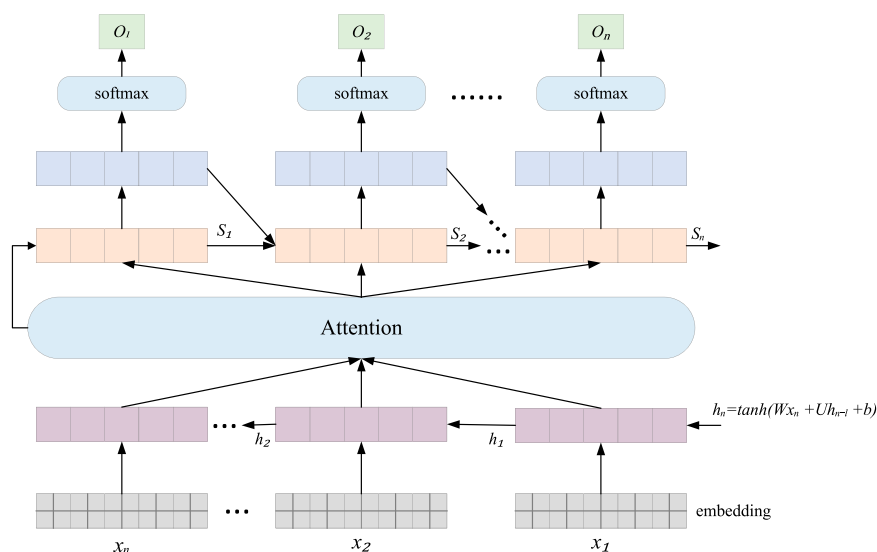


FIGURE 4
Flow chart of the Attention model.

between load data. This specific treatment of topology helps the model more accurately capture the energy consumption patterns of different types of loads, providing a stronger basis for prediction tasks.

The operation process of GCN Model is shown in Figure 2.

The main formula of GCN Model is as follows:

$$H^{(l+1)} = \sigma \left(\widehat{D}^{-\frac{1}{2}} \widehat{A} \widehat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (1)$$

Here, $H^{(l)}$ Represents the node feature matrix for layer l . σ Denotes the activation function, typically using ReLU, etc. \widehat{A} Indicates the symmetrically normalized adjacency matrix. \widehat{D} Represents the diagonal matrix of node degrees. $W^{(l)}$ Stands for the weight matrix for layer l .

In this formula, GCN gradually updates the feature representation of nodes through a multi-layer information transfer process, so that each node contains information about its surrounding nodes, thereby taking into account the influence of topological relationships. In the overall model, the role of GCN is to encode the topological structure information of the power system into a more information-rich feature representation, providing important basic information for subsequent load type prediction. Through the use of GCN, the model can better understand the relationship between nodes in the power system and improve the modeling ability of load type prediction problems. This is of great significance for comprehensively considering the complexity and diversity of the power system, thereby improving the accuracy of prediction and the ability to capture global information.

3.3 Sequence-to-sequence model

In our model, the Seq2Seq model (Sequence-to-Sequence model) is a key component for processing time series data and load type forecasting tasks Xiong et al. (2021). The basic principle of the Seq2Seq model is to map the input temporal sequence to

the output sequence through an encoder-decoder structure, while retaining and delivering key contextual information Takiddin et al. (2022). The role of the Seq2Seq model in the overall model approach is to take historical time series load data as the input sequence, and then generate the corresponding load type prediction sequence. The key to this process is to encode the rich information of the timing data into a fixed-length vector representation, which is then passed through a decoder to generate a sequence of load types. The encoder of the Seq2Seq model can effectively capture patterns and trends in historical time series data, while the decoder converts this information into load-type predictions Le et al. (2021). The encoder of the Seq2Seq model has excellent capabilities and can effectively capture patterns and trends in historical time series data. By learning representations of historical load data, the encoder is able to extract key temporal features, allowing the model to better understand the information required for load type forecasting tasks. This feature encoding method helps capture the complex relationships between load data, making the model more flexible and accurate when processing time series information. On the other hand, the decoder of the Seq2Seq model is able to effectively utilize the contextual information passed by the encoder when generating load type prediction sequences. By incorporating historical timing correlations into the generation process, the decoder is able to more accurately predict future load types. This end-to-end sequence modeling approach enables the model to perform well in load type prediction tasks, with higher accuracy and robustness compared to traditional methods and single deep learning models.

The operation process of Seq2Seq model is shown in Figure 3.

The main formula of Seq2Seq Model is as follows:

$$h_t = \text{Encoder}(x_t, h_{t-1}) \quad (2)$$

$$y_t = \text{Decoder}(h_t, y_{t-1}) \quad (3)$$

Here, h_t represents the hidden state of the encoder, which captures the information in the input sequence x_t and passes it to

TABLE 1 The comparison of different models in different indicators comes from National Grid Electricity Load Dataset, Canadian Electricity Load Dataset, U.S. Electricity Load Dataset and International Electricity Load Dataset.

Model	Datasets											
	National grid electricity load dataset				Canadian electricity load dataset				U.S. Electricity load dataset			
	Accuracy	Recall	F1 score	AUC	Accuracy	Recall	F1 score	AUC	Accuracy	Recall	F1 score	AUC
Wang et al. (2020)	85.37	88.53	85.77	92.21	89.61	91.31	85.47	89.62	96.18	83.14	81.27	87.18
Mohammadi (2021)	84.43	90.54	88.01	91.33	90.49	84.77	83.87	91.92	86.83	90.44	83.13	86.15
Alotaibi et al. (2020)	86.27	84.74	82.42	83.91	89.11	87.83	93.21	88.75	96.45	85.19	90.91	89.19
Alladi et al. (2019)	86.09	91.27	87.89	92.73	88.63	89.49	89.06	84.51	91.60	89.51	85.19	90.82
Hui et al. (2020)	87.41	89.38	83.71	91.15	87.40	92.31	90.78	93.04	88.87	87.79	85.75	87.34
Al-Badi et al. (2020)	86.26	91.46	81.13	94.41	96.10	90.63	85.07	91.81	87.90	92.89	85.08	89.92
Ours	96.22	93.54	91.06	94.45	97.11	94.53	91.46	95.17	95.63	97.55	94.87	97.37
									96.24	97.43	94.84	98.46

the decoder. y_t represents the output of the decoder, which is the predicted result of the load type. x_t represents the time series data for each time step of the input sequence. h_{t-1} and y_{t-1} represent the encoder hidden state and decoder output of the previous time step, respectively, for context transfer.

The encoder of the Seq2Seq model gradually encodes the historical time series data into hidden states h_t , and passes these hidden states to the decoder, which generates a sequence of load type predictions based on the hidden states. This process allows the model to make accurate load type predictions based on historical data and contextual information. The application of this model in this study plays a key role in helping the model better understand time series data, thereby improving the accuracy of load type prediction and global information capture capabilities.

3.4 Attention mechanism

In our model, the attention mechanism is a key component used to enhance modeling of the relationship between time series data and graph data Li et al. (2022c). The basic principle of this mechanism is to introduce a weight allocation mechanism in the encoder-decoder structure so that the model can focus on the information most relevant to the current prediction when generating load type predictions Massaoudi et al. (2021). In the overall model, the role of the attention mechanism is to enable the model to perform information fusion and selection based on the importance of different input data, thereby improving the accuracy of load type prediction. This mechanism dynamically adjusts the weight of the encoder output through the learned weight, allowing the model to more effectively capture the relationship between time series data and graph data, helping to improve prediction performance Zhang et al. (2020). The advantage of the attention mechanism is that it allows the model to be more flexible and intelligent when processing complex time series data and graph data. By introducing a weight allocation mechanism, the model is able to selectively focus on the part of the historical data that is relevant to the current prediction when predicting the load type at each time point. This dynamic adjustment feature enables the model to better adapt to changes in data distribution at different time points, improving the modeling capabilities of time series and graph data. In addition, the application of attention mechanism helps to improve the model's understanding of the complex topology of the power system, making it more sensitive to capture the correlation between nodes. In models that incorporate attention mechanisms, more targeted attention to key information helps optimize load type prediction performance.

The operation process of Attention Mechanism is shown in Figure 4.

The main formula of Attention Mechanism is as follows:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^T \exp(e_{ik})}$$

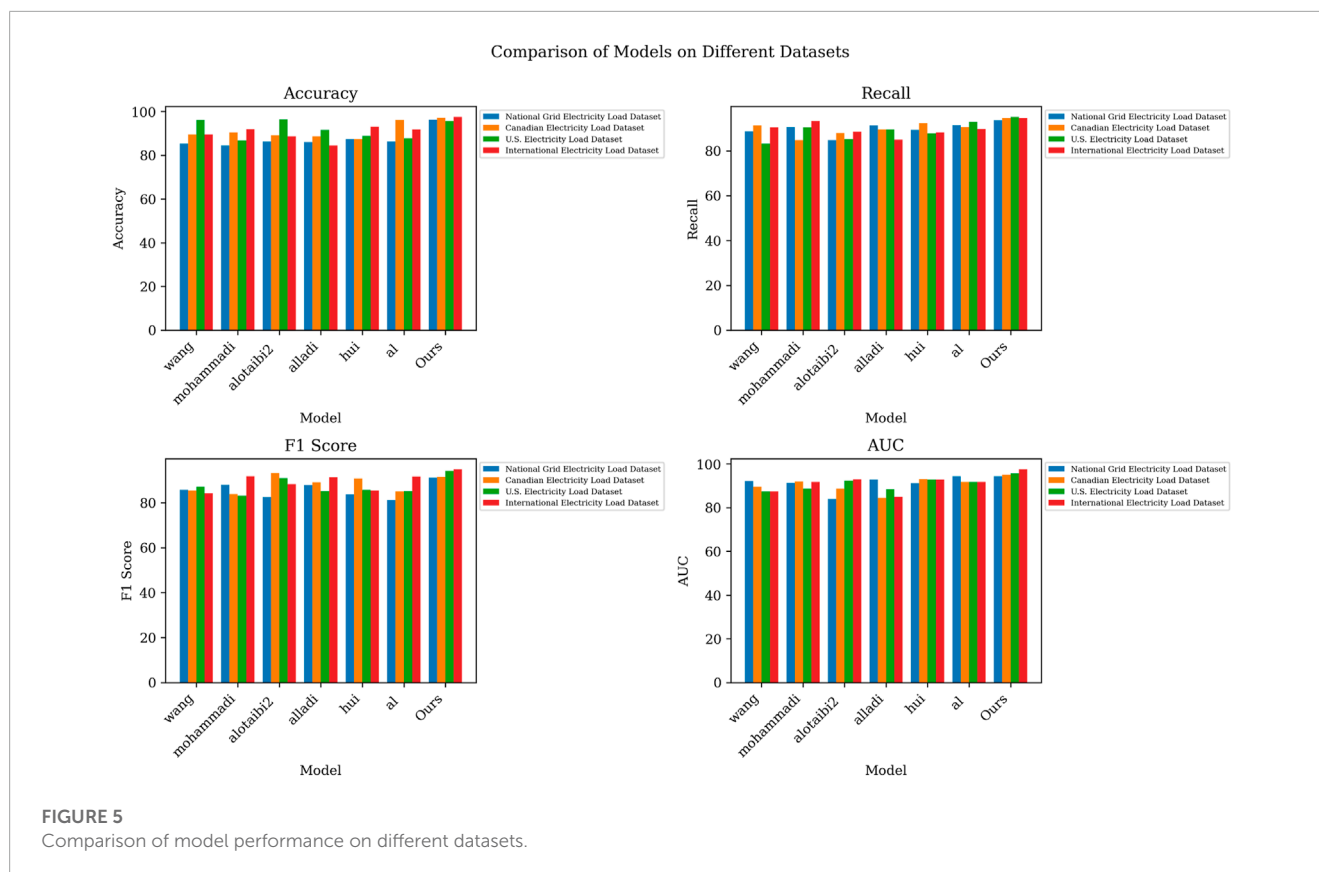
(4)

$$c_i = \sum_{j=1}^T \alpha_{ij} \cdot h_j$$

(5)

$$a_t = \text{Attention}(h_p, c_t)$$

(6)



Here, Q represents the attention weight of time step Q to time step Q , which is used to measure the importance of different time steps in time series data. Q represents the score for calculating the attention weight, usually obtained using inner product or other methods. Q represents the context vector at time step Q , which is obtained by weighted summation of the encoder output Q according to the attention weights. Q represents the output after applying attention, which is used for load type prediction.

The formulation of the attention mechanism describes how to calculate attention weights, context vectors, and apply attention to improve load type prediction. This mechanism plays a key role in the entire model and helps the model better understand and utilize the correlation between input data.

4 Experiment

4.1 Experimental environment

- Hardware Environment

The hardware environment used in the experiments consists of a high-performance computing server equipped with an AMD Ryzen Threadripper 3990X @ 3.70 GHz CPU and 1TB RAM, along with 6 Nvidia GeForce RTX 3090 24 GB GPUs. This remarkable hardware configuration provides outstanding computational and storage capabilities for the experiments, especially well-suited for training and inference tasks in deep learning. It effectively accelerates the

model training process, ensuring efficient experimentation and rapid convergence.

- Software Environment

In this study, we utilized Python and PyTorch to implement our research work. Python, serving as the primary programming language, provided us with a flexible development environment. PyTorch, as the main deep learning framework, offered powerful tools for model construction and training. Leveraging PyTorch's computational capabilities and automatic differentiation functionality, we were able to efficiently develop, optimize, and train our models, thereby achieving better results in the experiments.

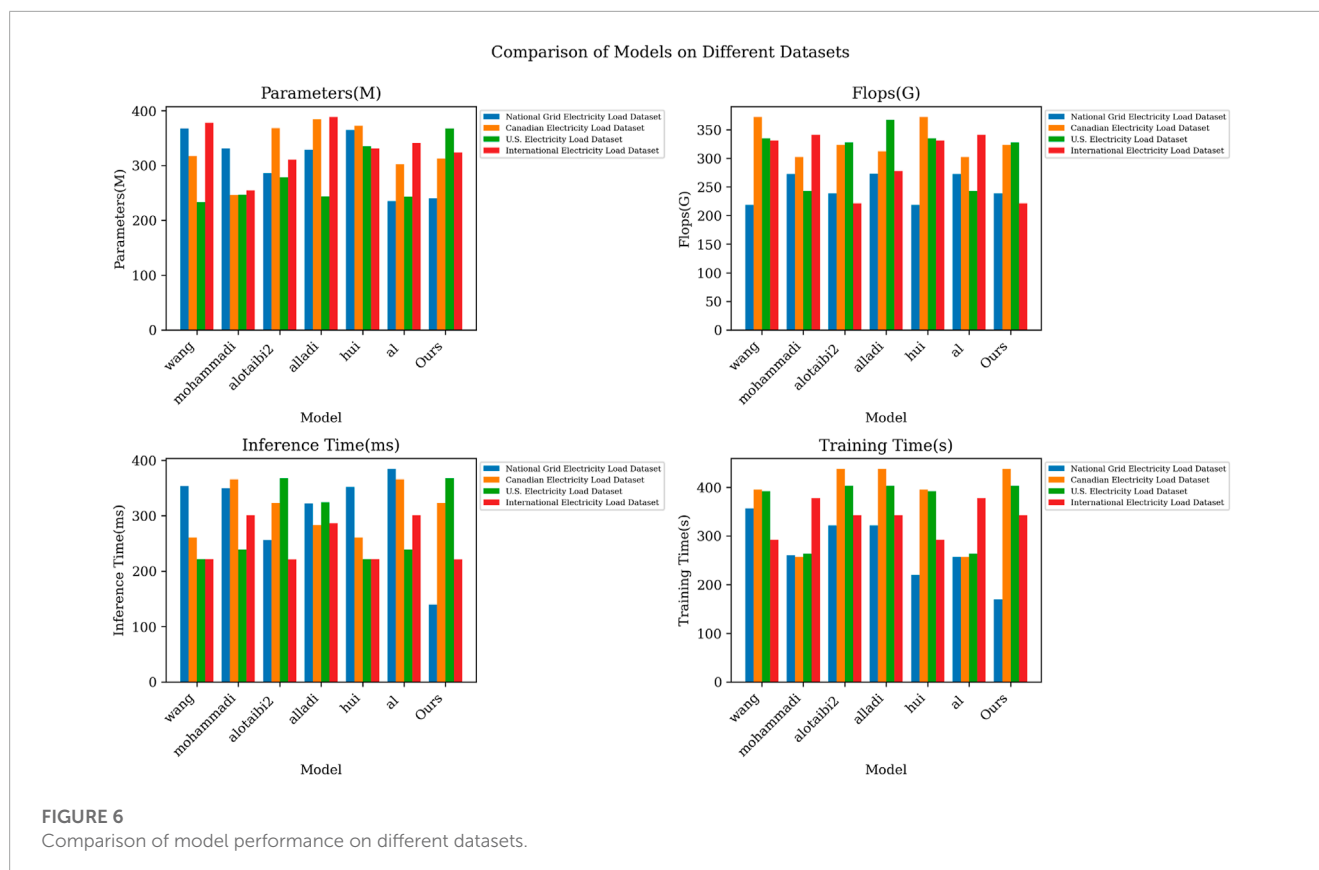
4.2 Experimental datasets

This paper mainly uses the following four data sets to study the problem of smart grid power load type prediction.

National Grid Electricity Load Dataset is a very important data set that provides key information for electric load forecasting research. The source of this data set is the State Grid of China, the largest domestic electricity supplier and operator in China. Data is carefully collected and maintained to ensure accuracy and reliability [Zhang and Hong \(2019\)](#). The data set includes multiple years of history, ranging from the past few years up to the most recent electricity load data. This long time span of data allows researchers to analyze seasonal and cyclical changes in electrical loads. The dataset covers different regions within China,

TABLE 2 The comparison of different models in different indicators comes from National Grid Electricity Load Dataset, Canadian Electricity Load Dataset, U.S. Electricity Load Dataset and International Electricity Load Dataset.

Method	Dataset															
	National grid electricity load dataset				Canadian electricity load dataset				U.S. Electricity load dataset				International electricity load dataset			
	Parameters (M)	Flops (G)	Inference Time (ms)	Training Time(s)	Parameters (M)	Flops (G)	Inferer Time (ms)	Training Time(s)	Parameters (M)	Flops (G)	Inference Time (ms)	Training Time(s)	Parameter (M)	Flops (G)	Inference Time (ms)	Training Time(s)
Wang et al. (2020)	367.46	218.53	353.56	356.53	331.13	272.29	349.21	260.46	286.15	238.48	226.97	328.77	272.92	255.75	321.74	321.48
Mohammadi (2021)	317.32	372.23	260.66	395.46	246.31	302.46	365.15	256.97	368.13	268.62	315.85	384.36	312.48	323.29	282.64	437.58
Alotaibi et al. (2020)	233.02	335.08	221.69	391.96	246.57	243.24	238.71	263.57	278.69	301.23	274.23	243.44	328.15	367.46	323.94	403.16
Alladi et al. (2019)	377.90	331.40	221.68	292.55	254.69	341.02	300.53	378.02	311.13	352.42	336.24	388.45	254.93	221.21	284.30	342.21
Hui et al. (2020)	364.86	216.82	351.79	219.93	235.27	205.14	384.18	254.49	284.02	393.12	334.41	218.72	277.72	291.13	286.52	361.66
Al-Badi et al. (2020)	352.07	374.63	221.36	318.28	308.49	214.32	395.20	346.63	224.08	391.79	226.52	300.86	343.67	293.43	388.01	207.65
Ours	155.22	131.23	228.77	241.43	240.17	213.87	139.59	159.33	156.57	169.47	197.89	157.64	231.01	219.01	215.05	235.17



including urban and rural areas. This covers China's wide range of geographical and climatic conditions, providing diversity for research. The importance of the National Grid Electricity Load Dataset cannot be underestimated. As data from the State Grid of China, it provides an opportunity to gain in-depth understanding of China's power system operations and load changes. This dataset is critical for power load type forecasting research as it contains rich information that helps researchers understand load patterns in different regions and seasons. In addition, as one of the world's largest electricity consumers, research on China's power system is of great significance to global power management and sustainable development.

Canadian Electricity Load Dataset is an important data resource that provides key information for electricity load forecasting studies. Sources for this data set include the Canadian government and electric utilities across Canada. These agencies are responsible for collecting and maintaining electrical load data to ensure data accuracy and availability. The Canadian Electricity Load Dataset covers multiple years of history, including the past few years up to the latest electrical load data. This long time span of data allows researchers to analyze seasonal and cyclical changes in electricity loads, as well as their evolution over time (Iqbal et al. (2021)). The dataset covers every province and city in Canada, including places with different climates and electricity needs. Due to Canada's geographical differences and climate diversity, this dataset is diverse and covers electricity load conditions under different conditions. Canadian Electricity Load Dataset is important in the study of electric load type forecasting. First, Canada is a geographically vast country with a variety of climatic and

topographic conditions, so this dataset provides information on electricity load characteristics under different meteorological and geographical conditions. Second, this dataset reflects the operation of the Canadian power system, which is critical for power load management and power system optimization. Most importantly, as a developed country, Canada's power system is modern and complex, so the study of power load type forecasting problems has special value.

U.S. Electricity Load Dataset is an important data resource that provides key information for electric load forecasting research. Sources for this data set include the U.S. Energy Information Administration (EIA) and various U.S. power companies (Lv et al. (2021)). These agencies collect and maintain electrical load data to ensure data accuracy and availability. The U.S. Electricity Load Dataset covers many years of history, ranging from the past few years up to the latest electricity load data. This long time span of data allows researchers to analyze seasonal and cyclical changes in electricity loads, as well as their evolution over time. The dataset covers every state and city in the United States, including places with different climates and electricity needs. As a country with geographical diversity and variable climate, the United States has diverse power load data, covering power load conditions under different conditions. The U.S. Electricity Load Dataset is important in power load type forecasting research, providing information on power load characteristics under different meteorological and geographical conditions, reflecting the dynamics of large-scale power supply and demand.

International Electricity Load Dataset brings together data from the International Energy Agency (IEA) and electricity companies

TABLE 3 Ablation experiments on the GCN-Seq2Seq module comes from National Grid Electricity Load Dataset, Canadian Electricity Load Dataset, U.S. Electricity Load Dataset and International Electricity Load Dataset.

Model	Datasets											
	National grid electricity load dataset				Canadian electricity load dataset				U.S. Electricity load dataset			
	Accuracy	Recall	F1 score	AUC	Accuracy	Recall	F1 score	AUC	Accuracy	Recall	F1 score	AUC
RNN	86.65	89.36	85.04	88.27	90.66	89.17	86.61	91.78	95.78	85.53	86.28	86.5
Resnet50	93.54	92.81	89.93	85.55	90.44	85.62	84.66	94.02	95.12	86.25	89.35	89.89
Resnet18	89.57	94.27	87.25	87.93	88.83	91.38	90.02	90.14	94.35	93.78	86.15	93.43
GCN-Seq2Seq	97.48	93.62	93.82	93.61	97.91	94.88	94.75	95.59	98.21	95.39	93.59	92.78
									97.95	95.67	94.5	96.24

in various countries and regions. The IEA is responsible for coordinating and collecting electricity load data in various countries to ensure the accuracy and availability of data. It covers many years of history, from the past few years up to the latest electrical load data. This long time span of data allows researchers to analyze seasonal and cyclical changes in electricity load, as well as electricity load trends on a global scale [Ahmad et al. \(2020\)](#). The dataset has a global geographical scope, covering multiple countries and regions. This makes it a diverse and comprehensive data resource, including places with different climates, cultures and power system characteristics. International Electricity Load Dataset is important in electric load type forecasting research. First, it reflects the operation of power systems in different countries and regions, providing key information for power load management and optimization on a global scale. Secondly, because it covers multiple countries and regions, this data set helps study cross-border power load forecasting problems and promotes international cooperation and knowledge sharing.

4.3 Experimental setup and details

This study uses the GCN-Seq2Seq model integrated with the attention mechanism to study the problem of smart grid power load type prediction. To ensure accuracy and reproducibility, experimental details need to be carefully designed. The experimental setup and details are as follows:

Step 1: Dataset preparation.

- Data sources: The four data sets come from the State Grid of China, the Canadian government and power companies, the U.S. Energy Information Administration (EIA), and the International Energy Agency (IEA). These datasets are historical power load information collected from different power systems.
- Time span: The data set covers many years of historical data, ranging from a few years to a few decades, to ensure that power load data under a variety of seasons and meteorological conditions are included.
- Geographic scope: These data sets cover different geographical scopes, including various regions in China, different regions in Canada, states and cities in the United States, as well as electricity load data on a global scale.
- Data cleaning and preprocessing: Before using the data, data cleaning and preprocessing are required, including removing missing values, processing outliers, data standardization, etc., to ensure the quality and consistency of the data.
- Data set division: The data set will be divided into a training set, a validation set and a test set. Usually 70% of the data is used for training, 15% is used for validation, and 15% is used for testing. This helps evaluate the performance and generalization ability of the model.

Step 2: Model selection and hyperparameter tuning.

- Model selection: We will consider using GCN, Seq2Seq, and overall models that introduce attention mechanisms. These

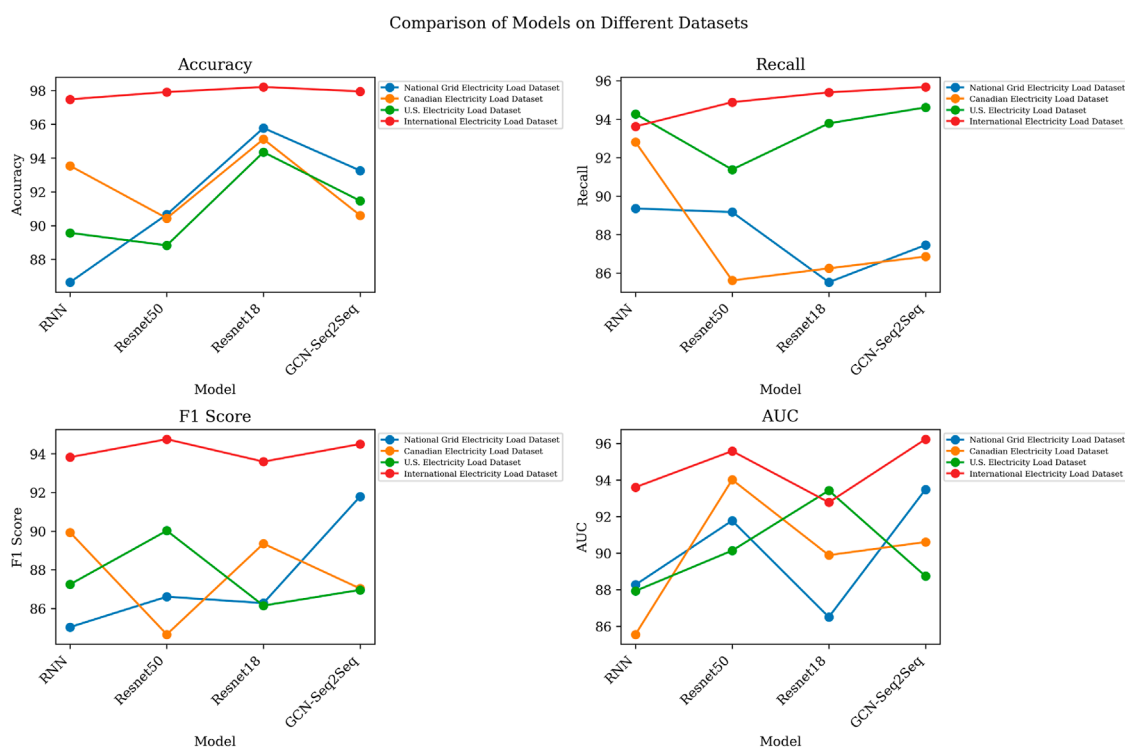


FIGURE 7

Comparison of model performance on different datasets.

models were chosen because of their advantages in processing graph data and time series data.

- Hyperparameter adjustment: In the experiment, we will perform hyperparameter adjustment, including the selection of key parameters such as learning rate, batch size, hidden layer size, and attention weight. We will use cross-validation to evaluate the performance of different hyperparameter settings.

Step 3: Model training process.

- GCN model training: For the GCN model, we will build the graph structure of the power system and use the adjacency matrix for training. GCN will utilize node features and graph structure information for training.
- Seq2Seq model training: For the Seq2Seq model, we will prepare time series data, including historical power load data as the input sequence, and load type as the output sequence. The Seq2Seq model will be trained using an encoder-decoder structure to learn load-type patterns.
- Holistic model training: In the holistic model, we will consider both the graph structure and the time series data of the power system. Attention mechanism will be used to capture the relationship between them. The overall model will be trained taking both data into account.

Step 4: Loss function and evaluation metrics.

- Loss function: We will choose an appropriate loss function to measure the performance of the model, depending on the nature of the problem. For classification tasks, the categorical cross-entropy loss function or the mean square error loss function is usually chosen.
- Evaluation metrics: We will use a series of evaluation metrics to measure the performance of the model, including accuracy, precision, recall, F1 score, etc. These metrics will be used for performance evaluation on the validation and test sets.

Step 5: Experimental Design.

- Ablation experiments: We will conduct ablation experiments to gradually evaluate the impact of each component of the model on overall performance. For example, we will study how the model performs without using the attention mechanism.
- Comparative experiments: We will conduct comparative experiments to compare and analyze our model with other commonly used deep learning models (such as CNN, RNN, TCN, GRU, DRL) to determine the superiority of our model.

Step 6: Results Analysis and Visualization.

- We will conduct a detailed analysis of the experimental results, comparing the performance of different models, the impact of hyperparameter settings, and performance on different data sets. We will use visualization tools to present key results to help gain insight into the model's behavior.

Comparison of Methods on Different Datasets (Parameters)

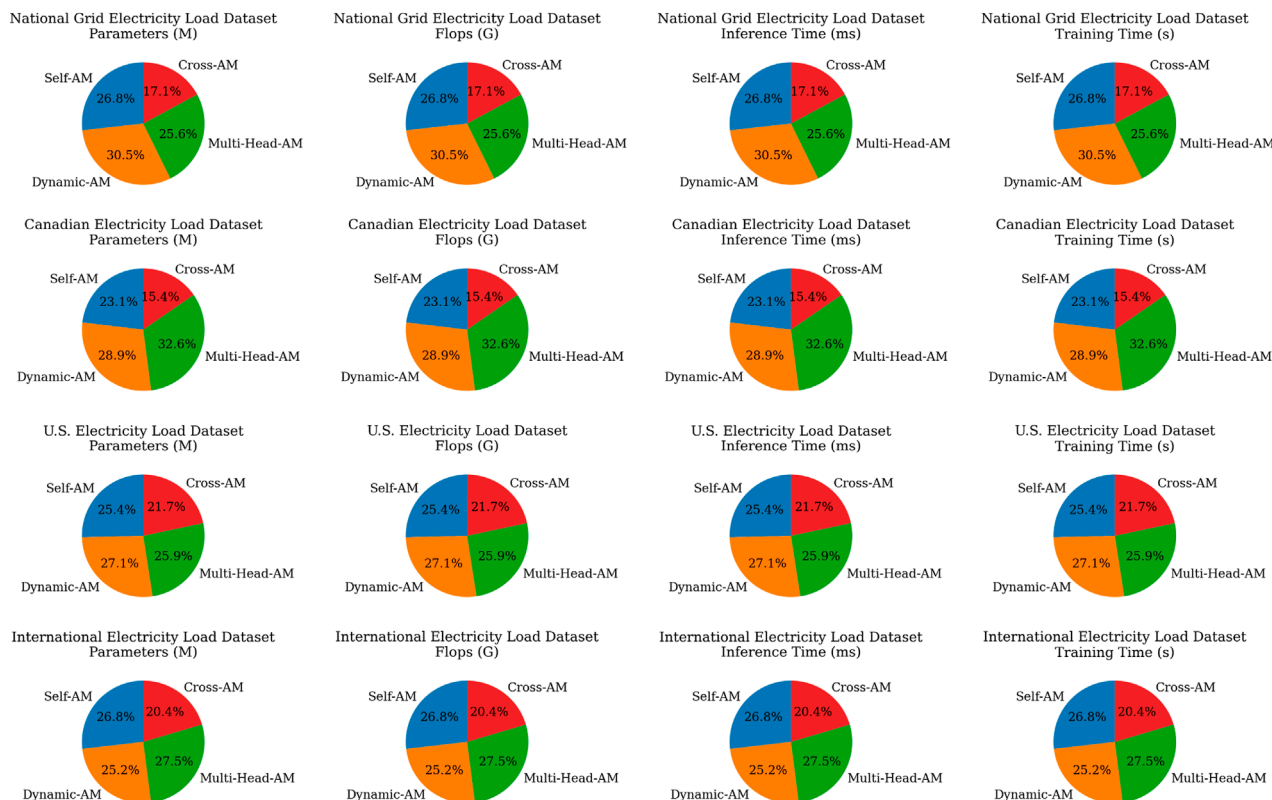


FIGURE 8

Comparison of model performance on different datasets.

4.4 Experimental results and analysis

During the experiment, we collected data including National Grid Electricity Load Dataset, Canadian Electricity Load Dataset, U.S. Electricity Load Dataset, International Electricity Load Dataset. Through experiments, we obtained the following results.

When we look at the results in Table 1, we can clearly see that our model performs significantly better than other models on different datasets. Specifically, on the National Grid Electricity Load Dataset, our model achieves 96.22% accuracy, 93.54% recall, 91.06% F1 score, and 94.45% AUC, which performance metrics significantly exceed other models, such as wang, mohammadi, alotaibi2, alladi and hui. On the Canadian Electricity Load Dataset, U.S. Electricity Load Dataset and International Electricity Load Dataset, our model also achieves the highest level of performance indicators, indicating its strong generalization ability on different data sets. Digging further into Figure 5, we can see that after visualizing the results from Table 1, the comparison of model performance becomes clearer. In this visualization, our model sits at the top of each dataset by a clear margin, outperforming other models. This visualization presents the superior performance of our model on different datasets, further confirming the excellent performance of our method in

power load type forecasting tasks. It should be emphasized that on the International Electricity Load Dataset, our model performed particularly well, reaching an AUC of 98.46%, which is much higher than other models. This shows that the introduction of the attention mechanism has important advantages for processing international-scale power load data and can more accurately capture the complex patterns of load types.

By analyzing the data in Table 2, we can clearly see the performance of our model on different data sets. First, we note that our model has a much lower number of model parameters than other models on each dataset. For example, on the National Grid Electricity Load Dataset, our model parameters are only 155.22M, while the number of parameters of other models exceeds 230M, which indicates that our model has a more lightweight design. Furthermore, our model has the lowest Flops and inference time on all datasets, further demonstrating its efficiency. This is critical due to resource constraints and response time requirements in real-world applications. After visualizing these performance indicators, as shown in Figure 6, we can see that our model achieves the best performance on each data set, which further confirms its superior effect in power load type forecasting tasks. It is worth noting that despite having fewer model parameters, our model performs particularly well on the International Electricity Load Dataset,

TABLE 4 Ablation experiments on the Cross Transformer module using different datasets.

Method	Dataset															
	National grid electricity load dataset				Canadian electricity load dataset				U.S. Electricity load dataset				International electricity load dataset			
	Parameters (M)	Flops (G)	Inference time (ms)	Training time (s)	Parameters (M)	Flops (G)	Infer time (ms)	Training time (s)	Parameters (M)	Flops (G)	Inference time (ms)	Training time (s)	Parameter (M)	Flops (G)	Inference time (ms)	Training time (s)
Self-AM	336.01	251.09	237.01	310.23	347.66	341.35	247.57	391.58	355.73	293.54	298.49	385.64	273.67	245.87	354.32	380.89
Dynamic-AM	383.13	314.22	252.49	291.59	327.17	357.61	365.72	350.76	361.81	265.79	242.14	283.27	373.96	293.19	226.47	386.15
Multi-Head-AM	320.99	353.55	241.41	318.34	343.52	331.01	230.23	365.21	325.45	306.16	229.15	284.00	351.07	285.38	378.83	395.05
Cross-AM	214.96	166.91	202.23	236.12	156.41	178.81	189.85	108.81	118.44	116.06	224.99	187.49	218.82	212.00	232.96	189.94

further verifying its generalization ability on different data sets. This shows that our model not only performs well in performance but also has a lightweight design that is applicable to various power load data sets.

By analyzing the data in Table 3, we can gain an in-depth understanding of the performance of the GCN-Seq2Seq module on different data sets and its impact on the overall performance of the model. First, we focus on the key performance indicators of the model on four different data sets, including accuracy (Accuracy), recall rate (Recall), F1 score (F1 Score) and AUC value (Area Under the Curve). On the National Grid Electricity Load Dataset, the GCN-Seq2Seq module achieved excellent performance, with an accuracy of 97.48%, a recall of 93.62%, an F1 score of 93.82%, and an AUC value of 93.61, significantly better than other models (RNN, Resnet50 and Resnet18). This shows that the GCN-Seq2Seq module has excellent classification performance in the power load type prediction task. On other data sets, the GCN-Seq2Seq module also performed well and maintained a high level of performance. Especially on the Canadian Electricity Load Dataset and International Electricity Load Dataset, the model's accuracy exceeded 97.9%, the recall rate exceeded 94.75%, the F1 score exceeded 94.5%, and the AUC values exceeded 95.59% and 96.24%. This further verifies the generalization ability and stability of the GCN-Seq2Seq module. After visualizing these performance indicators, as shown in Figure 7, we can clearly observe the excellent performance of the GCN-Seq2Seq module on different data sets, as well as its advantages over other models. The introduction of this surface attention mechanism module significantly improves the model's performance in power load type prediction tasks.

By analyzing the data in Table 4, we can gain an in-depth understanding of the performance of the Cross Transformer module on different data sets and its impact on the overall performance of the model. This table provides key performance indicators on four different data sets, including model parameters (Parameters), number of floating point operations (Flops), inference time (Inference Time) and training time (Training Time). First, let's focus on the performance of the Cross Transformer module on the National Grid Electricity Load Dataset. This module has a parameter volume of 214.96M, a floating point operation count of 166.91G, an inference time of 202.23 ms, and a training time of 236.12s. These metrics show the module's performance level when processing this data set. Then, we observe the performance of the Cross Transformer module on the other three datasets. On the Canadian Electricity Load Dataset, U.S. Electricity Load Dataset and International Electricity Load Dataset, the module has performance indicators of 156.41M, 178.81G, 189.85 ms and 108.81s respectively, and corresponding results of 118.44M, 116.06G, 224.99 ms and 187.49s numerical value. These data show the performance changes of the Cross Transformer module on different data sets. By visualizing these performance metrics, we can more clearly observe the performance of the Cross Transformer module on different data sets. As shown in Figure 8, the module performs poorly on the National Grid Electricity Load Dataset but has better performance on the other three datasets. This shows that the Cross Transformer module has certain flexibility and adaptability when dealing with different data distributions and tasks.

5 Conclusion and discussion

In this study, we focus on solving the problem of power load type prediction in smart grids to help the power system better understand and manage load changes. We propose an innovative deep learning model that combines graph convolutional network (GCN), sequence-to-sequence (Seq2Seq) model and attention mechanism to comprehensively consider the complex topology and time series data of the power system to achieve more accurate Load type forecasting. Specifically, we first use the GCN encoder to process the topological structure information of the power system and represent the node features into encoding of graph data. Next, the Seq2Seq decoder takes the historical time series data as the input sequence and generates a prediction sequence of the load type. In this process, an attention mechanism is introduced, allowing the model to fuse information based on the importance of different input data. Finally, the outputs of the GCN encoder and Seq2Seq decoder are integrated to achieve more accurate load type prediction. Through extensive experimental verification, we demonstrate the excellent performance of this model in load type forecasting tasks, significantly improving the accuracy of load type prediction in power systems.

Despite its remarkable results, this study suffers from two major flaws. First, the performance of our model in handling extreme situations needs to be further improved, such as sudden power load fluctuations, which require more robust processing capabilities. Secondly, our study still needs to be verified in more actual power systems to further confirm its generalization ability and robustness. Future research directions will consider improving the robustness of the model and extending the scope of experimental validation to more comprehensively evaluate its performance. It is also expected to explore more smart grid application areas, such as automated operation and maintenance of power systems and smart energy interaction, to further promote the development and application of smart grids.

This research provides an innovative method to solve the problem of power load type prediction and has important practical significance. By combining graph neural networks, sequence generation models, and attention mechanisms, we achieve more accurate predictions of power system load types, helping smart grids achieve more efficient energy management and optimization. This is of great significance to the high reliability, efficiency and sustainability of the power system, and also makes a positive contribution to the development of smart grids and sustainable energy integration.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

HS: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Writing–original draft. YR: Investigation, Methodology, Project administration, Resources, Writing–original draft. SW: Conceptualization, Formal Analysis, Methodology, Project administration, Resources, Software, Writing–original draft, Writing–review and editing. BZ: Investigation, Methodology, Project administration, Resources, Writing–review and editing. RY: Investigation, Project administration, Supervision, Visualization, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is supported by the Sub-project “Research on parameter design method of power electronic equipment control system based on new topology structure” of the Long Term Key Project of China Electric Power Research Institute, titled “New Power Electronic Control Technology and Equipment Prototype Supporting Safe and Efficient Operation of Large Power Grid” (Project No. XT83-23-007).

Acknowledgments

We would like to express our deep appreciation to the China Electric Power Research Institute for their generous support and funding, which made this research possible. We are also grateful to the Long Term Key Project for providing the framework and resources for our work.

Conflict of interest

Authors YR, SW, BZ, and RY were employed by China Electric Power Research Institute Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Ahmad, W., Ayub, N., Ali, T., Irfan, M., Awais, M., Shiraz, M., et al. (2020). Towards short term electricity load forecasting using improved support vector machine and extreme learning machine. *Energies* 13, 2907. doi:10.3390/en13112907
- Al-Badi, A. H., Ahshan, R., Hosseinzadeh, N., Ghorbani, R., and Hossain, E. (2020). Survey of smart grid concepts and technological demonstrations worldwide emphasizing on the Oman perspective. *Appl. Syst. Innov.* 3, 5. doi:10.3390/asi3010005
- Alladi, T., Chamola, V., Rodrigues, J. J., and Kozlov, S. A. (2019). Blockchain in smart grids: a review on different use cases. *Sensors* 19, 4862. doi:10.3390/s19224862
- Alotaibi, I., Abido, M. A., Khalid, M., and Savkin, A. V. (2020). A comprehensive review of recent advances in smart grids: a sustainable future with renewable energy resources. *Energies* 13, 6269. doi:10.3390/en13236269
- Arumugham, V., Ghanimi, H. M., Pustokhin, D. A., Pustokhina, I. V., Ponnamp, V. S., Alharbi, M., et al. (2023). An artificial-intelligence-based renewable energy prediction program for demand-side management in smart grids. *Sustainability* 15, 5453. doi:10.3390/su15065453
- Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., et al. (2021). Cnn variants for computer vision: history, architecture, application, challenges and future scope. *Electronics* 10, 2470. doi:10.3390/electronics10202470
- Chen, R., Wang, Y., Li, G., Yan, D., and Cao, H. (2022). "Pre-training models based knowledge graph multi-hop reasoning for smart grid technology," in *Proceedings of 2021 5th Chinese conference on swarm intelligence and cooperative control* (Springer), 1866–1875.
- Cheon, H., Dziewulska, K. H., Moosic, K. B., Olson, K. C., Gru, A. A., Feith, D. J., et al. (2020). Advances in the diagnosis and treatment of large granular lymphocytic leukemia. *Curr. Hematol. malignancy Rep.* 15, 103–112. doi:10.1007/s11899-020-00565-6
- Daniels, J., Doukas, P. G., Escala, M. E. M., Ringbloom, K. G., Shih, D. J., Yang, J., et al. (2020). Cellular origins and genetic landscape of cutaneous gamma delta t cell lymphomas. *Nat. Commun.* 11, 1806. doi:10.1038/s41467-020-15572-7
- Dhruv, P., and Naskar, S. (2020). Image classification using convolutional neural network (cnn) and recurrent neural network (rnn): a review. *Mach. Learn. Inf. Process. Proc. ICMLIP* 2019, 367–381. doi:10.1007/978-981-15-1884-3_34
- Fan, J., Zhang, K., Huang, Y., Zhu, Y., and Chen, B. (2023). Parallel spatio-temporal attention-based tcn for multivariate time series prediction. *Neural Comput. Appl.* 35, 13109–13118. doi:10.1007/s00521-021-05958-z
- Hafeez, G., Alimgeer, K. S., and Khan, I. (2020). Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Appl. Energy* 269, 114915. doi:10.1016/j.apenergy.2020.114915
- Han, S.-Y., Zhao, Q., Sun, Q.-W., Zhou, J., and Chen, Y.-H. (2022). Engs-dgr: traffic flow forecasting with indefinite forecasting interval by ensemble gc, seq2seq, and dynamic graph reconfiguration. *Appl. Sci.* 12, 2890. doi:10.3390/app12062890
- Hossain, M. J., and Rahnamay-Naeini, M. (2021). "State estimation in smart grids using temporal graph convolution networks," in *Proceedings of the 2021 north American power symposium (NAPS)*, College Station, TX, USA, November 2021 (IEEE), 01–05.
- Huang, X., Hong, S. H., Yu, M., Ding, Y., and Jiang, J. (2019). Demand response management for industrial facilities: a deep reinforcement learning approach. *IEEE Access* 7, 82194–82205. doi:10.1109/access.2019.2924030
- Hui, H., Ding, Y., Shi, Q., Li, F., Song, Y., and Yan, J. (2020). 5g network-based internet of things for demand response in smart grid: a survey on application potential. *Appl. Energy* 257, 113972. doi:10.1016/j.apenergy.2019.113972
- Ibrahim, M. S., Dong, W., and Yang, Q. (2020). Machine learning driven smart electric power systems: current trends and new perspectives. *Appl. Energy* 272, 115237. doi:10.1016/j.apenergy.2020.115237
- Iqbal, H. K., Malik, F. H., Muhammad, A., Qureshi, M. A., Abbasi, M. N., and Chishty, A. R. (2021). A critical review of state-of-the-art non-intrusive load monitoring datasets. *Electr. Power Syst. Res.* 192, 106921. doi:10.1016/j.epsr.2020.106921
- Le, T.-T.-H., Heo, S., and Kim, H. (2021). Toward load identification based on the hilbert transform and sequence to sequence long short-term memory. *IEEE Trans. Smart Grid* 12, 3252–3264. doi:10.1109/tsg.2021.3066570
- Leng, J., Ruan, G., Song, Y., Liu, Q., Fu, Y., Ding, K., et al. (2021). A loosely-coupled deep reinforcement learning approach for order acceptance decision of mass-individualized printed circuit board manufacturing in industry 4.0. *J. Clean. Prod.* 280, 124405. doi:10.1016/j.jclepro.2020.124405
- Li, Q., Zhu, Y., Ding, J., Li, W., Sun, W., and Ding, L. (2022a). Deep reinforcement learning based resource allocation for cloud edge collaboration fault detection in smart grid. *CSEE J. Power Energy Syst.* doi:10.17775/CSEEJPES.2021.02390
- Li, Y., Nie, J., and Chao, X. (2020). Do we really need deep cnn for plant diseases identification? *Comput. Electron. Agric.* 178, 105803. doi:10.1016/j.compag.2020.105803
- Li, Y., Ren, R., Huang, B., Wang, R., Sun, Q., Gao, D. W., et al. (2022b). Distributed hybrid-triggering-based secure dispatch approach for smart grid against dos attacks. *IEEE Trans. Syst. Man, Cybern. Syst.* 53, 3574–3587. doi:10.1109/tsmc.2022.3228780
- Li, Y., Wei, X., Li, Y., Dong, Z., and Shahidepour, M. (2022c). Detection of false data injection attacks in smart grid: a secure federated deep learning approach. *IEEE Trans. Smart Grid* 13, 4862–4872. doi:10.1109/tsg.2022.3204796
- Liu, L.-N., Yang, G.-H., and Wasly, S. (2023). Distributed predefined-time dual-mode energy management for a microgrid over event-triggered communication. *IEEE Trans. Industrial Inf.* 1–11. doi:10.1109/tii.2023.3304025
- Lv, L., Wu, Z., Zhang, J., Zhang, L., Tan, Z., and Tian, Z. (2021). A vmd and lstm based hybrid model of load forecasting for power grid security. *IEEE Trans. Industrial Inf.* 18, 6474–6482. doi:10.1109/tii.2021.3130237
- Massaoudi, M., Abu-Rub, H., Refaat, S. S., Chihi, I., and Oueslati, F. S. (2021). Deep learning in smart grid technology: a review of recent advancements and future prospects. *IEEE Access* 9, 54558–54578. doi:10.1109/access.2021.3071269
- Mohammadi, F. (2021). Emerging challenges in smart grid cybersecurity enhancement: a review. *Energies* 14, 1380. doi:10.3390/en14051380
- O'Dwyer, E., Pan, I., Acha, S., and Shah, N. (2019). Smart energy systems for sustainable smart cities: current developments, trends and future directions. *Appl. Energy* 237, 581–597. doi:10.1016/j.apenergy.2019.01.024
- Peng, S., Zhang, Z., Deng, R., and Cheng, P. (2023). Localizing false data injection attacks in smart grid: a spectrum-based neural network approach. *IEEE Trans. Smart Grid* 14, 4827–4838. doi:10.1109/tsg.2023.3261970
- Reddy, N. P., Pasdeloup, D., Zadeh, M. K., and Skjetne, R. (2019). "An intelligent power and energy management system for fuel cell/battery hybrid electric vehicle using reinforcement learning," in *Proceedings of the 2019 IEEE transportation electrification conference and expo (ITEC)*, Detroit, MI, USA, June 2019 (IEEE), 1–6.
- Takiddin, A., Ismail, M., Zafar, U., and Serpedin, E. (2022). Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids. *IEEE Syst. J.* 16, 4106–4117. doi:10.1109/jsyst.2021.3136683
- Wang, X., Liu, Y., and Choo, K.-K. R. (2020). Fault-tolerant multisubset aggregation scheme for smart grid. *IEEE Trans. Industrial Inf.* 17, 4065–4072. doi:10.1109/tii.2020.3014401
- Xiao, J., and Zhou, Z. (2020). "Research progress of rnn language model," in *Proceedings of the 2020 IEEE international conference on artificial intelligence and computer applications (ICAICA)*, Dalian, China, June 2020 (IEEE), 1285–1288.
- Xiong, G., Przystupa, K., Teng, Y., Xue, W., Huan, W., Feng, Z., et al. (2021). Online measurement error detection for the electronic transformer in a smart grid. *Energies* 14, 3551. doi:10.3390/en14123551
- Xu, A., Wu, T., Zhang, Y., Hu, Z., and Jiang, Y. (2021). "Graph-based time series edge anomaly detection in smart grid," in *Proceedings of the 2021 7th IEEE intl conference on big data security on cloud (BigDataSecurity)*, IEEE intl conference on high performance and smart computing, (HPSC) and IEEE intl conference on intelligent data and security (IDS), NY, USA, May 2021 (IEEE), 1–6.
- Ye, Y., Qiu, D., Wu, X., Strbac, G., and Ward, J. (2020). Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning. *IEEE Trans. Smart Grid* 11, 3068–3082. doi:10.1109/tsg.2020.2976771
- Zhang, F., Liu, Q., Liu, Y., Tong, N., Chen, S., and Zhang, C. (2020). Novel fault location method for power systems based on attention mechanism and double structure gru neural network. *IEEE Access* 8, 75237–75248. doi:10.1109/access.2020.2988909
- Zhang, Z., and Hong, W.-C. (2019). Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. *Nonlinear Dyn.* 98, 1107–1136. doi:10.1007/s11071-019-05252-7
- Zhang, Z., Zhang, D., and Qiu, R. C. (2019). Deep reinforcement learning for power system applications: an overview. *CSEE J. Power Energy Syst.* 6, 213–225. doi:10.17775/CSEEJPES.2019.00920
- Zhou, S., Hu, Z., Gu, W., Jiang, M., Chen, M., Hong, Q., et al. (2020). Combined heat and power system intelligent economic dispatch: a deep reinforcement learning approach. *Int. J. Electr. power and energy Syst.* 120, 106016. doi:10.1016/j.ijepes.2020.106016



OPEN ACCESS

EDITED BY

Hengrui Ma,
Qinghai University, China

REVIEWED BY

Yixin Liu,
Tianjin University, China
Jun Li,
Nanjing Institute of Technology (NJIT),
China
Chenhao Sun,
Changsha University of Science and
Technology, China

*CORRESPONDENCE

Zhenbing Zhao,
✉ zhaozhenbing@ncepu.edu.cn

RECEIVED 19 October 2023

ACCEPTED 19 December 2023

PUBLISHED 08 January 2024

CITATION

Sun S, Guo W, Wang Q, Tao P, Li G and
Zhao Z (2024), Optimal scheduling of
microgrids considering real power losses
of grid-connected microgrid systems.
Front. Energy Res. 11:1324232.
doi: 10.3389/fenrg.2023.1324232

COPYRIGHT

© 2024 Sun, Guo, Wang, Tao, Li and
Zhao. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Optimal scheduling of microgrids considering real power losses of grid-connected microgrid systems

Shengbo Sun¹, Wei Guo¹, Qiwei Wang², Peng Tao², Gang Li² and
Zhenbing Zhao^{2*}

¹State Grid Hebei Marketing Service Center, Shijiazhuang, Hebei, China, ²North China Electric Power University, Baoding, Hebei, China

Energy conservation, emission reduction and vigorous development of new energy are inevitable trends in the development of the power industry, but factors such as energy storage loss, solar energy loss and line loss in real power situations have led the problem to a complex direction. To address these intricacies, we use a more precise modeling approach of power loss and propose a collaborative optimization method integrating the Deep-Q-Network (DQN) algorithm with the multi-head attention mechanism. This algorithm calculates weighted features of the system's states to compute the Q-values and priorities for determining the next operational directives of the energy system. Through extensive simulations that replicate real world microgrid (MG) scenarios, our investigation substantiates that the optimization methodology presented here effectively governs the distribution of energy resources. It accomplishes this while accommodating uncertainty-induced losses, ultimately achieving the economic optimization of MG. This research provides a new approach to deal with problems such as energy loss, which is expected to improve economic efficiency and sustainability in areas such as microgrids.

KEYWORDS

microgrid, energy management, deep reinforcement learning (deep RL), real power loss, attention mechanism (AM)

1 Introduction

1.1 Background and related works

With the exacerbating energy crisis and environmental pollution, solar and wind energy have played an increasingly vital role as distributed energy resources due to their abundant and pollution-free nature. However, solar and wind energy are random and intermittent, posing difficulties for grid integration and dispatch. Microgrids have emerged as an effective solution to facilitate the comprehensive utilization of renewable energy (Zhang and Kang, 2022). Microgrids show enormous potential in resolving renewable energy integration thanks to their flexible operation and ease of control. Their efficient and cost-effective operation is a prerequisite for sustainable development. Nevertheless, the multi-source characteristic of renewable energy sources introduces complexity to the control problem in microgrid systems. Based on recent surveys, it has been observed that as much as 13% of the total generated power is dissipated as losses at the distribution level (Wu et al., 2010; Patel

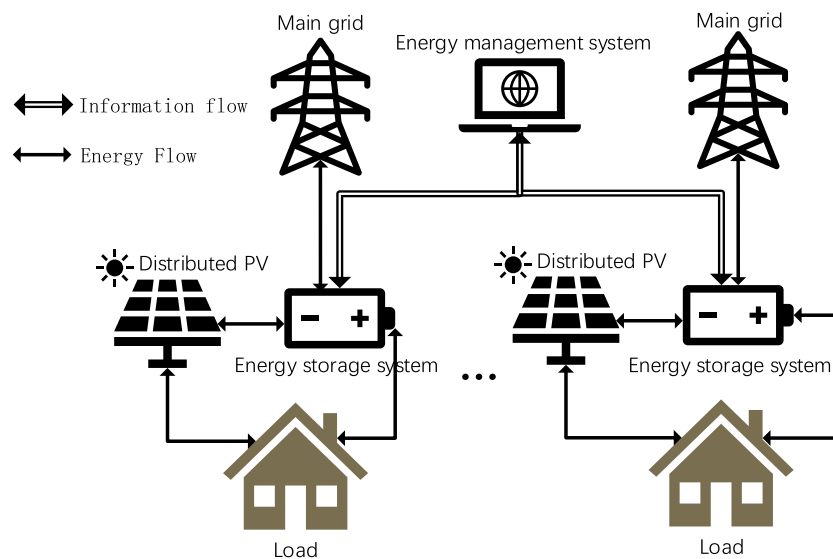


FIGURE 1
MG system structure diagram.

and Patel, 2016) applied ant colony optimization (ACO) to the reconfiguration of microgrids with distributed generation (DG) in order to minimize power losses (Kumari et al., 2017). introduced a particle swarm optimization (PSO) approach aimed at reducing DG costs and enhancing the voltage profile while addressing power loss concerns. Both of the aforementioned methods ascertain the optimal placement of DG using optimization algorithms. However, they do not account for the distinction between linear and nonlinear loads in their calculations. On the basis of this problem, this paper proposes a more accurate model of the actual line loss.

Energy system scheduling for microgrids has been investigated in a number of previous studies. Numerous studies utilize model-based control paradigms, including model predictive control (MPC) (Gan et al., 2020), mixed-integer linear programming (MILP) (Paterakis et al., 2015), dynamic and stochastic programming (Farzaneh et al., 2019), and alternating direction method of multiplier (ADMM) (Ma et al., 2018). However, once a large number of DERs connected to the MG in a disorderly way, the operation of the power grid will be largely influenced by its randomness and uncertainty. This makes it difficult to obtain the accurate system model. To solve these challenges, a model-free technique using reinforcement learning (RL) has been proven beneficial for energy system scheduling since the model of the environment is not necessary in this method. It is now emerging as the pre-eminent tool for unknown environmental decision-making issues. The authors of (Kim et al., 2016) present an RL algorithm that enables service providers and customers to acquire pricing and energy consumption strategies without any prior knowledge, thus reducing system costs (Fang et al., 2020). explored a dynamic RL-based pricing scheme to attain optimal prices when dealing with fast-charging electric vehicles connected to the grid. To reduce the electricity bills of residential consumers, a model for load scheduling using RL was developed in the literature (Lee and Choi, 2022), where the residential load

includes dispatches-available load, non-dispatches-available load, and local PV generation. In recent research findings, to address the dynamically changing operational conditions of appliances, a federated DQN approach has been proposed for managing energy in multiple homes (Remani et al., 2019). This research showcased exceptional performance of the DQN method in addressing continuous state space energy management challenges. Nevertheless, in MG scenarios, the performance of the DQN model in energy scheduling is significantly compromised by the inherent uncertainty of renewable energy sources. Furthermore, there is currently no well-defined strategy in place to address the complex issue of multivariate losses.

1.2 Contributions

To overcome the aforementioned challenges, this paper proposes an optimization method for grid-connected MG energy storage scheduling based on the DQN cooperative algorithm, aiming at minimizing the cost of electricity expenses, which is named AP DQN. Specifically, the proposed algorithm combines the multi-headed attention mechanism with the PER mechanism in DQN to improve its performance. In this configuration, the DQN interacts with the environment to obtain Q values and form rewards, and uses prioritized experience replay to stabilize learning. In addition, the algorithm computes the weighted features of the state using the multi-headed attention mechanism, and uses the weighted features to compute the Q-value and priority, which can make the state-action pair information of the terminal closer to the merit-seeking target, thus improving the overall convergence speed of the DQN. The case study verifies the effectiveness of the proposed algorithm for grid-connected MG energy storage scheduling with real-world data. The MPCLP algorithm is subsequently benchmarked against the optimal global solution.

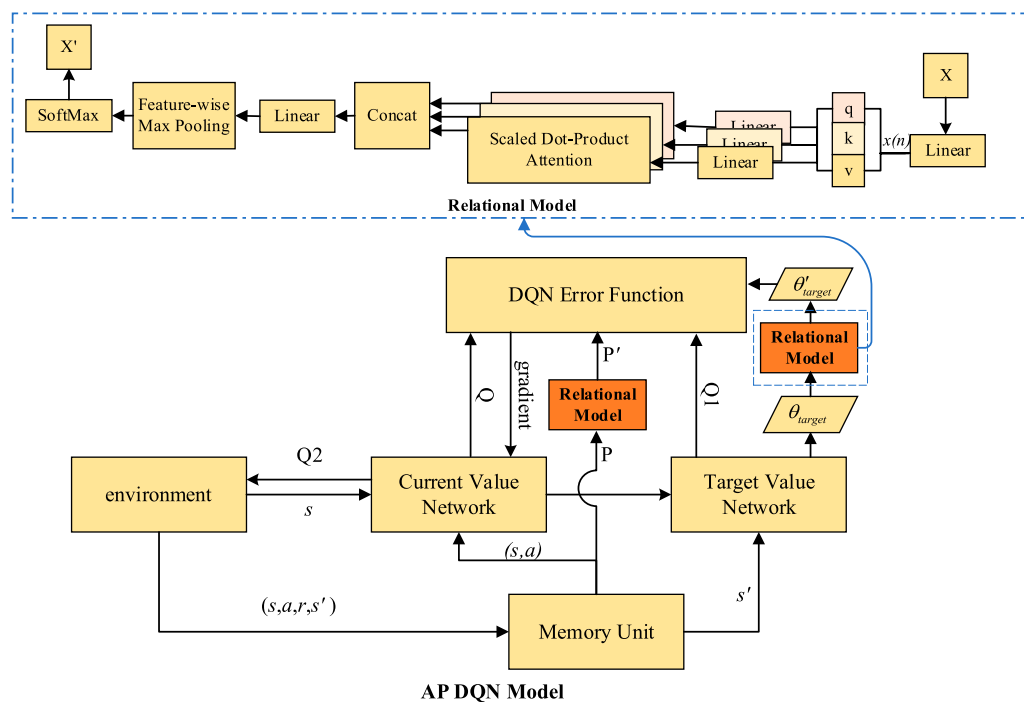


FIGURE 2
Algorithm operation process diagram.

TABLE 1 Hyper-parameters.

ϵ_{str}	ϵ_{stp}	d	err	α	β
1.1	0.01	0.0001	0.01	0.8	0.6

The primary contributions of this paper can be summarized as follows:

1. A precise mathematical model encompassing both linear and nonlinear power losses is developed to address the issue of multivariate loss factors in MGs.
2. A game combination optimization scheme based on deep reinforcement learning algorithm DQN is constructed based on the problem of difficult to handle multivariate uncertainties in MGs.
3. The AP DQN algorithm incorporating the multi-head attention mechanism is proposed for the problem of lossy features. Experimental results show that the method greatly improves the exploration efficiency. From the perspective of cost objective, our model outperforms the standard DQN by 33.5% and outperforms the MPCLP-based mechanism by up to 17.74%.

2 Microgrid's DRL model

2.1 Environment model

The environment model serves as the MG system environment that interacts with the agent. In this project, we considered a MG

with internal user loads, a photovoltaic field and an energy storage system (ESS), which is connected to the main grid through only one distribution line. Figure 1 illustrates the conceptual MG model that is envisioned in this study. The MG is managed by an energy management system (EMS), which fully controls all operations of the MG, including the processes of charging and discharging the ESS, as well as the power trading activities between the MG and the main grid. To enhance the stability and ensure the uninterrupted operation of mission-critical activities, it is necessary to monitor the state of the microgrid's emergency load reserve during main grid outages, called the state of charge (SOC) in the following article. We divide the MG system into 24 time slots and each time slot is denoted as t . To enable analytical calculations, the microgrid's power is assumed to be balanced, and a quasistatic time-varying energy model is employed.

Reinforcement learning can be characterized as a Markov Decision Process (MDP) comprising a state space \mathcal{S} , an action space \mathcal{A} , a utility or payoff function r (utility and payoff functions are used in the report), a state transfer probability matrix P and a discount factor γ (Moradi et al., 2018). The learning process is the process of making action decisions after obtaining the next state and reward return through the interaction between the agent and the environment, and then continuously optimizing. The discount factor γ modulates the agent's consideration of the long-term consequences of their decisions on future states: (1) small values of γ force agents to focus more on the immediate payoffs of the next few steps and significantly reduce the payoffs of future steps; (2) large values of γ force actors to think more strongly about future payoffs and thus become more farsighted.

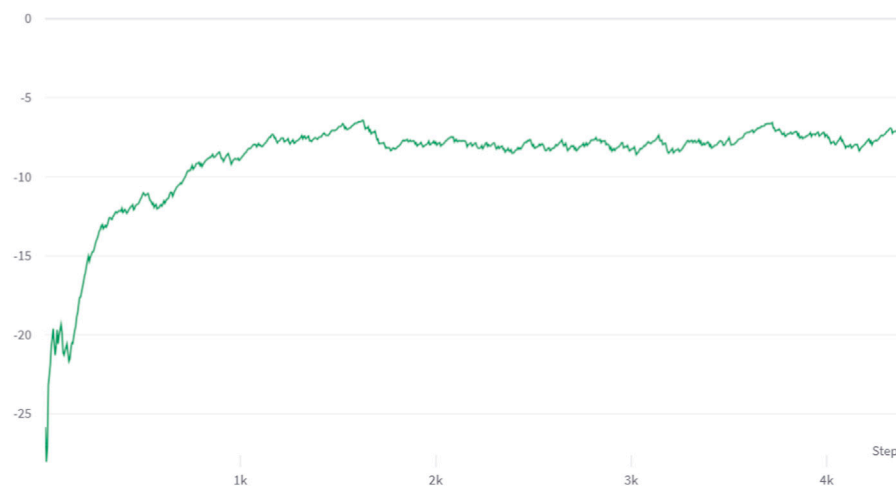


FIGURE 3
Mean episode reward with AP DQN (smoothing 0.8).

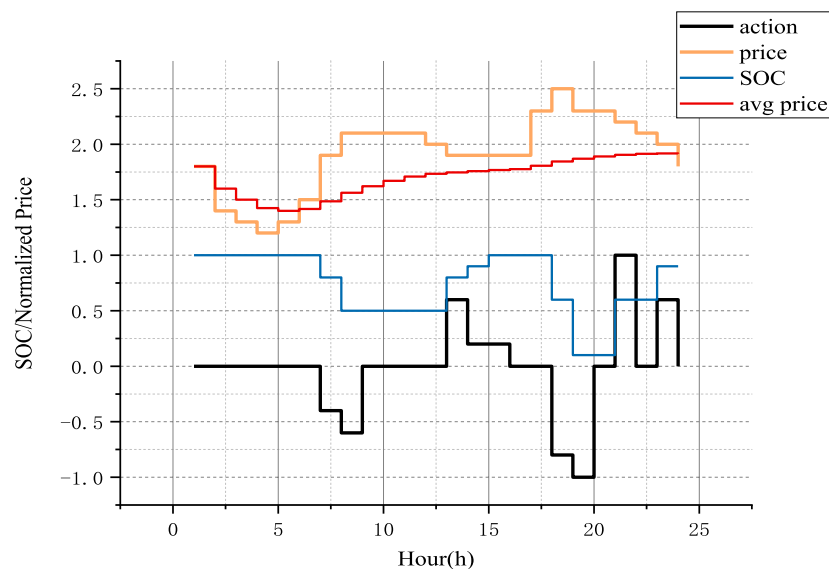


FIGURE 4
24-h average MG scheduling partial result with AP DQN.

2.1.1 ESS model

In this system, ESS mainly performs charging and discharging operations with an action space range of -1 to 1 . A positive value represents charging, while a negative value indicates discharging. We define $A_t \in \{-1, -0.8, \dots, 0.8, 1\}$ as the discrete action set. In each time slot t , the ESS is limited to performing either a charging action or a discharging action, but not both simultaneously. The state of the SOC is updated as follows (Chen and Su, 2018):

$$SOC_{t+1} = \begin{cases} SOC_t + \frac{A_t \times P_r \times \eta_c \times \Delta t}{E_r \times \eta_d}, & A_t \geq 0 \\ SOC_t + \frac{A_t \times P_r \times \Delta t}{E_r \times \eta_d}, & \text{else} \end{cases}$$

where parameters η_c, η_d, P_r, E_r represent the charging efficiency of the ESS, discharging efficiency of the ESS, rated power of the ESS, and energy storage capacity of the ESS, respectively. The energy trading mechanism incorporates the consideration of wear and tear costs. The ESS wear cost coefficient, denoted as k , is defined as follows:

$$k = \frac{C_i}{\eta_d \times E_r \times \delta \times N_c}$$

where parameters C_i, δ, N_c represent the initial investment cost of the ESS, the depth-of-discharge and the number of life cycles at a rated of the depth-of-discharge, respectively.

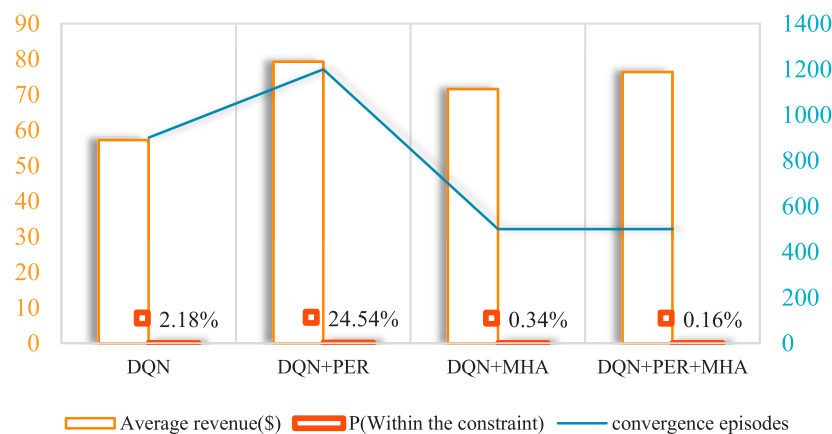


FIGURE 5
Results of ablation experiments.

2.1.2 PV model

The DC power generated by the PV module undergoes filtering in the DC circuit to eliminate current fluctuations and electromagnetic interference. It is then converted into AC power in the inverter circuit. The resulting AC power is rectified to obtain sinusoidal AC power. Subsequently, the output-side filter circuit is employed to mitigate high-frequency interference signals generated during the inverting process. This enables integration into the grid or direct supply to the load. The losses incurred during these transformations can be mathematically expressed as follows:

$$P_{loss}^{pv} = (P_{loss}^{DC} + P_{loss}^{AC}) / P_C$$

where parameters P_{loss}^{pv} , P_{loss}^{DC} , P_{loss}^{AC} , P_C represent the photovoltaic inverter losses, the DC/AC loss and the installed capacity, respectively.

2.2 Real power loss of loads

Given the diverse characteristics of loads and their varying operational conditions, we adopt distinct methods for evaluating power losses. In the case of linear loads, we calculate losses by subtracting the output power from the input power to achieve greater accuracy. For nonlinear loads, we consider power factor adjustments to account for the influence of factors such as harmonics and phase differences. The expression for real power loss in the load is as follows:

$$P_L = \sum_{i=1}^N [(P_i^{lin} - P_i^{out}) + \delta_p (\bar{P}_i^{nin} - \bar{P}_i^{nout})]$$

where parameters P_L , P_i^{lin} , P_i^{out} represent the real power loss of loads, the linear loads power input, the linear loads power output. The parameters δ_p , \bar{P}_i^{nin} , \bar{P}_i^{nout} represent the power factor, the average nonlinear loads power input, the average nonlinear loads power output (Sima et al., 2023). The N act as the number of loads.

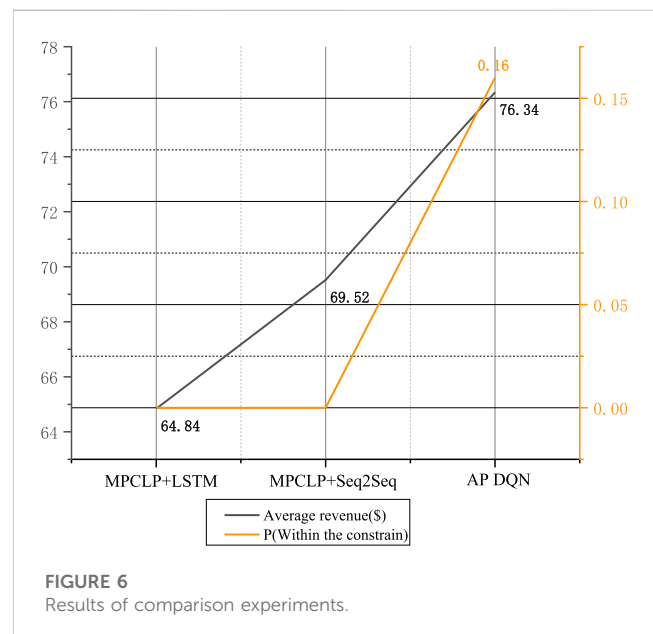


FIGURE 6
Results of comparison experiments.

2.3 Objective function and constrains designs

To keep the energy trading decisions of the MG within a reasonable range, we specify that the ESS must reserve enough energy for the critical tasks, named the target SOC, to minimize the MG operation cost under this constraint. The constraint functions are as follows:

$$A_t \times \frac{P_r}{E_r} \leq 1 - SOC_t, A_t \in (0, 1]$$

$$A_t \times \left(-\frac{P_r}{E_r} \right) \leq SOC_t - SOC_{target}, A_t \in [-1, 0]$$

With such a constraint, the system is able to reserve enough emergency power for the MG in the case of an accident scenario. In addition, the objective function of the optimization is described as follows:

TABLE 2 Comparison and ablation results of different model.

Technique	ρ (%)	Avg revenue (\$)	Avg convergence episodes
MPCLP + LSTM	0.00	64.84	—
MPCLP + Seq2Seq	0.00	69.52	—
DQN	2.18	57.18	9000
DQN + PER	24.54	79.28	12000
DQN + MHA	0.34	71.58	4000
Our Model	0.16	76.34	2000

The bold values represents the method we proposed.

$$Obj_t = \min \sum_{t=1}^{24} \left[(Pr_t + k) \times A_t \times P_r \times \frac{L_t}{L_t - P_L + PV_t'} \right]$$

in which $PV_t' = PV_t - P_{loss}^{pv}$

where Pr_t denotes the electric price at time t , L_t denotes the consumer load power at time t , and PV_t' represents the actual PV power in the MG.

3 Materials and methods

First, this paper designs a more accurate mathematical model of multivariate loss factors for microgrids with respect to loss uncertainty as well as ambiguity. Then, based on the problem of loss feature diversity, an optimization scheme of deep reinforcement learning algorithm DQN combined with multi-attention mechanism is proposed, which utilizes the principle of attention to process loss data of different sizes more efficiently, and ultimately derives the optimal scheduling actions of the energy management system for microgrids according to the objective of economic optimization.

3.1 Heading baseline-DQN

The issue examined in this paper pertains to a high-dimensional uncertainty problem that is not amenable to traditional algorithmic solutions. Reinforcement learning is a frontier area of machine learning and is a hot topic in the field of intelligent systems research. Reinforcement learning distinguishes itself from supervised learning in terms of the availability of training labels or targets. In supervised learning, the correct labels are provided to train the model. In contrast, reinforcement learning operates without explicit targets and adopts a trial-and-error approach. The model learns from its past mistakes to iteratively enhance its decision-making abilities for future actions (Mnih et al., 2013).

In the traditional approach to solving the reinforcement learning problem, a Q-table is constructed to store the Q-values, which represent the expected rewards of taking specific actions in particular states. The Q-table is updated utilizing an iterative updating rule that takes into account the recursive relationship between the Q-values. Nevertheless, when a continuous state space is encountered, it becomes impractical to create a state-action table to record every possible combination of

states and actions. To overcome the limitation, a neural network known as the DQN is employed. The DQN takes the states as inputs and generates the Q-values for each possible action as outputs, which is trained through the trial-and-error process (Mnih et al., 2013). The Q-values are subsequently updated using the Bellman equation as follows:

$$Q(S_t, A_t) = r_t + \gamma \times \max_{A_{t+1}} (Q(S_{t+1}, A_{t+1}))$$

where $Q(S_t, A_t)$ is the Q-value at time t , and $\max_{A_{t+1}} (Q(S_{t+1}, A_{t+1}))$ denotes the maximum Q-value taking optimal action at the subsequent step. Under the policy, the value of taking action A_t at S_t must equivalent to the expected reward of transitioning to the next state S_{t+1} plus the discounted expected Q-value of taking the best decision A_{t+1} at S_{t+1} (Mnih et al., 2013). The interdependence among the Q-values at consecutive steps ensures that the iterative update rule enables the discovery of an optimal policy, leading to the convergence of Q-values towards their optimal values. This recursive relationship facilitates the convergence of the Q-value iteration process, allowing for the determination of an optimal policy.

3.2 AP DQN method

In this section, we design the AP DQN method. There are two main modules in this algorithm, one of them is a learning network model based on PER DQN, and the other is a relational network model that includes the multi-head attention mechanism. The multi-head attention mechanism in our work is applied to focus on relevant samples in the experience replay process as well as the Q-value handling process. The innovations of this algorithm are mainly represented in the following: the multi-head attention mechanism is adopted to enable the network to process the input sequences in parallel, and the model is able to realize the information fusion and sharing so as to enhance the learning ability; the network structure is improved comprehensively, and the addition of the relational model layer to weight the Q-value provides stronger adaptive learning flexibility for the network weights. The algorithm operation process diagram is shown in Figure 2. In this certain case, we put the mathematical models of ESS and PV and constraints of devices in the environment module.

Where Q indicates $Q(S_t, A_t)$, $Q1$ indicates $\max_{A_{t+1}} (Q(S_{t+1}, A_{t+1}))$, and $Q2$ indicates $\arg\max_{A_{t+1}} (Q(S_{t+1}, A_{t+1}))$. p indicates the stored experience tuple (s, a, r, s') .

We use a multi-headed attention mechanism in experience replay memories. This allows for the selection of experience replay samples by using the multi-headed attention mechanism to focus on all past samples in the memory pool and select those that are most important and relevant for the current learning. At the same time, the multi-headed attention mechanism is used in the calculation of Q values. When calculating $Q(S_t, A_t)$ values for each action A_t under the state S_t , the different features of s can be weighted using multi-headed attention, so that the Q value calculation focuses more on those state features that are most important at the moment. This can produce more accurate Q-value estimates.

3.2.1 Algorithmic framework

In this section we design the algorithm framework, the operation process is as follows:

1. First, we initialize the playback memory unit, the priority weights array P , and the Q network and target network parameters.
2. We capture the experience tuple (s, a, r, s') in the environment and store it in the memory unit.
3. For each experience tuple (s, a, r, s') stored, the attention-weighted feature x' of s is computed using the multi-headed attention mechanism:
 - (a). Calculate attention headers with number of K :

$$\text{attn_head}_k = \text{Softmax}(W_k X_s + b_k), k = 1, 2, \dots, K$$

where W_k indicates the attention parameter matrix, X_s is the matrix corresponding to the state, and b_k is decided by the attention value.

- (b). Fuse the attention header to obtain the final attention value attn .
- (c). Calculate the weighted characteristics:

$$x' = \sum_{i=1}^N \text{attn}[i] * X_s[i]$$

4. Calculate the priority p of each tuple, Q_{target} is calculated using the target network parameters, and θ_{target} is the target network parameter:

$$p = (|r| + \gamma * \max_a (Q_{\text{target}}(x', a'; \theta_{\text{target}})))^\alpha$$

5. Select the experience tuple with the number of batch size for learning by priority.
 - (a). Calculate the Q value for each experience using the Q network and the weighted state x' .
 - (b). Calculate the loss of each experience using the Q target:

$$L = (Q_{\text{target}} - Q(x', a'; \theta))^2$$

- (c). Gradient descent updates the Q-network parameters θ .
6. Update the priority array P and the target network parameter θ_{target}
 7. Repeat steps 2-6 for training.

3.2.2 Reward function design

A segmented reward function is designed to guide the trading strategy provided that all conditions are satisfied, where the reward value depends on:

1. The state difference from the target SOC.
2. The final cost obtained from the MG operation.

Below the target SOC, it is imperative to prioritize charging the ESS promptly, irrespective of the price. Similarly, the price must be high enough to discharge the ESS below the target SOC. Therefore, the charging and discharging criteria for the ESS differ depending on whether the SOC is below or above the target level. To optimize the utilization of the remaining storage capacity, the charging price for the ESS should decrease as the state of charge (SOC) approaches full SOC. This incentivizes efficient charging when there is ample capacity available. Conversely, the price for discharging the ESS should be higher when the SOC is closer to the target SOC. This approach encourages the effective utilization of the remaining available energy and ensures that the SOC is maintained at the desired level. In addition to this setting, two penalty factors are introduced to have further control of the ESS operational behavior. The first penalty term Pnt_t^{ESS} is applied when the action chosen by the agent violates a constraint within the system. The second penalty term Pnt_t^{PV} is assigned when the ESS with available energy capacity fails to store excess solar energy. The first penalty term is introduced to account for the constraint of the ESS, aiming to extend the operational lifetime of the unit, while the second penalty term serves the purpose of maximizing the storage of solar energy within capacity limit of the ESS. The reward function R is as follows:

$$R(P_t, Pr_t^{\text{avg}}, \text{SOC}_t | A_t) = (Pr_t^{\text{avg}} - (Pr + k)) \times (\text{SOC}_{t+1} - \text{SOC}_t) \times E_r - Pnt_t^{\text{ESS}} - Pnt_t^{\text{PV}}$$

$$Pr_t^{\text{avg}} = \frac{\sum_{t=0}^{24} Pr_t}{24}$$

$$Pnt_t^{\text{ESS}} = \begin{cases} 0, & \text{else} \\ 10, & \text{if } \text{SOC}_t + A_t > 1 \text{ or } \text{SOC}_t - A_t < -1 \end{cases}$$

$$Pnt_t^{\text{PV}} = \begin{cases} 0, & \text{if } PV_t \leq (L_t + A_t \times P_r) \\ \exp(2.5 \times (1 - \text{SOC}_{t+1}))1 - \text{SOC}_{t+1}), & 2.5 \times -1, \text{ if } PV_t > (L_t + A_t \times P_r) \end{cases}$$

where Pr_t^{avg} represents the average price observed throughout the 24 time slots preceding time t .

3.2.3 Relational model

The main idea of the relational model is the weighted encoding of states using a multi-headed attention mechanism. The attention mechanism can be understood as a process of addressing information, where the attention value is computed by calculating the attention distribution based on the key and associating it with the value. This computation is performed with respect to a task-specific query vector Q , allowing the attention mechanism to focus on relevant information and selectively combine it with the query. By dividing each query, key, and value into multiple branches, multiple different attention calculations are performed on Q , K , and V to obtain multiple different outputs, and then these different outputs are stitched together to obtain the final output. Indeed, this process represents the essence of attention, which helps mitigate the complexity of neural networks. Instead of feeding all N inputs into the network for computation, attention selectively chooses task-relevant information to be inputted. This approach is similar

to the concept of gating mechanisms in Recurrent Neural Networks (RNNs), where the network learns to focus on relevant information and effectively allocate computational resources (Azam and Younis, 2021).

Due to the priority sampling strategy, PER introduces a bias towards selecting higher priority samples during training (Schaul et al., 2015). This bias has the potential to lead to overfitting of the results obtained by the DQN algorithm. Therefore, to correct for bias, we introduce the relational model to adjust the sampling weights. The built-in attention mechanism allows direct monitoring of the training process by highlighting the areas that agents focus on when making decisions. It naturally incorporates the policy gradient algorithm in reinforcement learning, where each time-step attention mechanism samples from $L = \mathbf{m}^* \mathbf{m}$ to a position requiring attention based on a random attention policy π_g . This policy is represented using a neural network whose output is composed of the probabilities of location selection. Among them, the formula for calculating and updating the policy gradient algorithm is as follows:

$$\begin{aligned}\nabla J(\theta) &= \sum_s \mu_\pi(s) \sum_a q_\pi(s, a) \nabla_\theta \pi(a|s, \theta) \\ &= E_\pi \left[\gamma^t \sum_a q_\pi(S_t, a) \nabla_\theta \pi(a|S_t, \theta) \right] \\ &= E_\pi \left[\gamma^t \sum_a q_\pi(S_t, a) \pi(a|S_t, \theta) \frac{\nabla_\theta \pi(a|S_t, \theta)}{\pi(a|S_t, \theta)} \right] \\ &= E_\pi \left[\gamma^t q_\pi(S_t, A_t) \frac{\nabla_\theta \pi(A_t|S_t, \theta)}{\pi(A_t|S_t, \theta)} \right] \\ &= E_\pi \left[\gamma^t G_t \frac{\nabla_\theta \pi(A_t|S_t, \theta)}{\pi(A_t|S_t, \theta)} \right] \\ \theta_{t+1} &= \theta_t + \alpha \gamma^t G_t \frac{\nabla_\theta \pi(A_t|S_t, \theta)}{\pi(A_t|S_t, \theta)}\end{aligned}$$

where $\nabla J(\theta)$ indicates the strategy gradient and G_t indicates the cumulative rewards. α indicates the step length and γ indicates the discount factor.

4 Experiments and results

In this section, we present simulation results to demonstrate the effectiveness of the proposed algorithm. These results serve as empirical evidence supporting the performance and efficacy of the algorithm. Specifically, the DQN architecture employed in this study consists of one input layer with four neurons, three fully connected hidden layers with 40 and 80 neurons, and one output layer with 14 neurons. This configuration allows for effective learning and decision-making within the energy management algorithm. ϵ greedy strategy and hyperparameters of PER are listed in Table 1. The mean episode reward with AP DQN is shown in Figure 3. The customer load, solar power and dynamic tariff are obtained from the self-built datasets. The P_r and E_r of the lithium-ion battery ESS used in the experiment are 1,000 kW and 5,000 kWh respectively.

Due to the large range of resultant data, we chose the average MG scheduling results over a time horizon of 24 h as a demonstration of the scheduling strategy, and the result with AP

DQN is shown in Figure 4. Due to the large time horizon involved in the dataset, the obtained ESS scheduling strategy is measured in terms of the final economic cost and the percentage of the system working within the constraints. We used a model predictive control linear programming (MPCLP) based algorithm (Matute et al., 2018) for comparison and performed ablation experiments. MPCLP is a linear programming optimization method, which commonly employs an optimization software to work out the problem. It provides good optimization accuracy while satisfying the assumptions of a linear dynamic system. Among them, MPCLP uses two prediction models, LSTM and Seq2Seq, respectively. The results of the ablation experiments are shown in Figure 5. The results of the comparison experiments are shown in Figure 6.

As seen in Figure 4, Positive values of action in the figure indicate charging, negative values indicate discharging, and SOC ranges from 0 to 1. It can be concluded that the EMS will combine the state of the SOC at the moment with the floating tariff to give the best possible action within the constraints.

As seen in Figure 5, the base DQN has poor performance in the ablation experiment, but the average gain rises significantly with the addition of PER, however, this is a result of large-scale constraint violations. With the addition of the multi-head attention mechanism, the algorithm is able to obtain an average return close to the PER DQN while maintaining a certain range of constraints. After adding the multi-headed attention mechanism to DNQ together with PER, the result of maximizing the average gain and minimizing the probability of constraint violation can be obtained.

As seen in Figure 6, AP DQN has the highest average profit in the comparison experiment, but there is a default rate of 0.16%, although this is an acceptable range. The reason for this is that the traditional linear programming approach has a strict adherence to the constraints and therefore a p -value of 0. In contrast, the proposed AP DQN algorithm can violate the constraints driven by the reward values to a minor degree, thus achieving the goal of maximizing the average profit.

The results obtained from the comparative experiments and ablation studies using different models are summarized in Table 2. Comparison and Ablation Results of Different Model. As can be seen from the table that our model outperforms the standard DQN by 33.5%, the MPCLP based mechanism by 17.74% at most. Compared with PER DQN, our model is a better choice in terms of algorithmic efficiency and conditional constraints.

5 Conclusion

In this paper, we propose an AP DQN algorithm. The algorithm not only maximizes monetary benefits but also maintains the reliability of the MG at the same time, being able to maintain sufficient energy reserves for critical operations. The algorithm presented uses a multi-headed attention mechanism as well as a prioritized experience replay mechanism to use current information for optimizing energy trading decisions. The algorithm we propose is a model-free reinforcement learning method, which usually has strong generalization ability. This method learns a wide range of strategies from a large amount of empirical data so that it can make reasonable decisions in uncovered states and can adapt better to various situations and conditions. In

comparison with the MPCLP approach, it can be concluded that the reinforcement learning based approach has a higher average monetary benefit in the presence of higher system reliability. It is worth noting that the reward function in RL can be further adjusted and optimized to improve the overall results. Fine-tuning the reward function has a significant impact on the performance of the RL algorithm. Additionally, it is important to consider that value-based RL methods generate discrete trading decisions, whereas MPCLP decisions are continuous in nature. This distinction can affect the comparison of results obtained from the two approaches. In future work, policy-based reinforcement learning is an appropriate direction to be investigated to obtain continuous decisions.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The dataset in this article is a power industry dataset that cannot be disclosed according to regulations. Requests to access these datasets should be directed to ZZ, zhaozhenbing@ncepu.edu.cn.

Author contributions

SS: Conceptualization, Formal Analysis, Investigation, Project administration, Supervision, Writing–original draft. WG: Data curation, Formal Analysis, Methodology, Writing–original draft. QW: Validation, Visualization, Writing–original draft. PT: Investigation, Validation, Writing–review and editing. GL: Formal Analysis, Investigation, Visualization, Writing–review and

editing. ZZ: Conceptualization, Data curation, Methodology, Supervision, Writing–original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the S&T Program of Hebei (22284504Z).

Acknowledgments

Heartfelt thanks to everyone who contributed to this paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Azam, M. F., and Younis, M. S. (2021). Multi-horizon electricity load and price forecasting using an interpretable multi-head self-attention and EEMD-based framework. *IEEE Access* 9, 85918–85932. doi:10.1109/ACCESS.2021.3086039
- Chen, T., and Su, W. (2018). Local energy trading behavior modeling with deep reinforcement learning. *IEEE Access* 6 (2), 62806–62814. doi:10.1109/ACCESS.2018.2876652
- Fang, C., Lu, H., Hong, Y., Liu, S., and Chang, J. (2020). Dynamic pricing for electric vehicle extreme fast charging. *IEEE Trans. Intell. Transp. Syst.* 22 (1), 531–541. doi:10.1109/TITS.2020.2983385
- Farzaneh, H., Shokri, M., Kebriaei, H., and Aminifar, F. (2019). Robust energy management of residential nanogrids via decentralized mean field control. *IEEE Trans. Sustain. Energy* 11 (3), 1995–2002. doi:10.1109/TSTE.2019.2949016
- Gan, L. K., Zhang, P., Lee, J., Osborne, M. A., and Howey, D. A. (2020). Data-driven energy management system with Gaussian process forecasting and mpc for interconnected microgrids. *IEEE Trans. Sustain. Energy* 12 (1), 695–704. doi:10.1109/TSTE.2020.3017224
- Kim, B. G., Zhang, Y., Schaar, M. V., and Lee, J. W. (2016). Dynamic pricing and energy consumption scheduling with reinforcement learning. *IEEE Trans. Smart Grid* 7 (5), 2187–2198. doi:10.1109/TSNG.2015.2495145
- Kumari, R. L., Kumar, G. N., Nagaraju, S. S., and Jain, M. B. (2017). "Optimal sizing of distributed generation using particle swarm optimization," in 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies, Kerala, India, July, 2017. doi:10.1109/ICICT1.2017.8342613
- Lee, S., and Choi, D. H. (2022). Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources. *IEEE Trans. Ind. Inf.* 18 (1), 488–497. doi:10.1109/TII.2020.3035451
- Ma, W. J., Wang, J., Gupta, V., and Chen, C. (2018). Distributed energy management for networked microgrids using online admm with regret. *IEEE Trans. Smart Grid* 9 (2), 847–856. doi:10.1109/TSNG.2016.2569604
- Matute, J. A., Marcano, M., Zubizarreta, A., et al. (2018). "Longitudinal model predictive control with comfortable speed planner," in IEEE International Conference on Autonomous Robot Systems and Competitions, Torres Vedras, Portugal, April, 2018, 25–27. doi:10.1109/ICARSC.2018.8374161
- Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2013). Playing atari with deep reinforcement learning. *CoRR*. doi:10.48550/arXiv.1312.5602
- Moradi, H., Esfahanian, M., Abtahi, A., and Zilouchian, A. (2018). Optimization and energy management of a standalone hybrid micro-grid in the presence of battery storage system. *Energy* 147, 226–238. doi:10.1016/j.energy.2018.01.016
- Patel, A. G., and Patel, C. (2016). "Distribution network reconfiguration for loss reduction," in 2016 International Conference on Electrical, Electronics, and Optimization Techniques, Chennai, India, March, 2016, 3937–3941. doi:10.1109/ICEEOT.2016.7755453
- Paterakis, N. G., Erdinc, O., Bakirtzis, A. G., and Catalao, J. P. S. (2015). Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies. *IEEE Trans. Ind. Inf.* 11 (6), 1509–1519. doi:10.1109/TII.2015.2438534
- Remani, T., Jasmin, E. A., and Ahamed, T. P. I. (2019). Residential load scheduling with renewable generation in the smart grid: a reinforcement learning approach. *IEEE Syst. J.* 13 (3), 3283–3294. doi:10.1109/JSYST.2018.2855689
- Schaul, T., Quan, J., Antonoglou, I., et al. (2015). Prioritized experience replay. *CoRR*. doi:10.48550/arXiv.1511.05952
- Sima, L., Miteva, N., and Dagan, K. J. (2023). A novel approach to power loss calculation for power transformers supplying nonlinear loads. *Electr. Power Syst. Res.* 223, 109582. doi:10.1016/j.epsr.2023.109582
- Wu, Y., Lee, C. Y., Liu, L. C., and Tsai, S. H. (2010). Study of reconfiguration for the distribution system with distributed generators. *IEEE Trans. Power Deliv.* 25 (3), 1678–1685. doi:10.1109/TPWRD.2010.2046339
- Zhang, Z., and Kang, C. (2022). Challenges and prospects for constructing the new-type power system towards a carbon neutrality future. *Proc. CSEE* 42, 2806–2819. doi:10.13334/j.0258-8013.pcsee.220467



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Guangsheng Pan,
Southeast University, China
Hongxia Wang,
University of Denver, United States

*CORRESPONDENCE

Jialun Sun,
✉ 202108580021223@ctgu.edu.cn

RECEIVED 29 May 2023

ACCEPTED 17 July 2023

PUBLISHED 08 January 2024

CITATION

Xiong C, Xu L, Ma L, Hu P, Ye Z and Sun J
(2024), Research on large-scale clean
energy optimal scheduling method based
on multi-source data-driven.
Front. Energy Res. 11:1230818.
doi: 10.3389/fenrg.2023.1230818

COPYRIGHT

© 2024 Xiong, Xu, Ma, Hu, Ye and Sun.
This is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Research on large-scale clean energy optimal scheduling method based on multi-source data-driven

Chuanyu Xiong¹, Lingfeng Xu², Li Ma¹, Pan Hu³, Ziyong Ye⁴ and Jialun Sun^{4*}

¹State Grid Hubei Electric Power Company Limited Economic Research Institute, Wuhan, Hubei, China, ²State Grid Hubei Electric Power Co., Ltd., Wuhan, Hubei, China, ³State Grid Hubei Electric Power Research Institute, Wuhan, Hubei, China, ⁴College of Electrical and New Energy Engineering, China Three Gorges University, Yichang, Hubei, China

With the large-scale growth and grid connection of intermittent renewable energy such as wind and solar, the problem of increasing renewable energy curtailment rate and system backup flexibility has become increasingly prominent. In order to solve the problem of high proportion of renewable energy scientific consumption and flexible and stable operation of energy system. We propose a flexible and economical dispatch method based on data-driven multi-regional power system. For the problem of economic dispatch of multi-area power system, a mathematical calculation model is established to satisfy the constraints of unit output, system power balance, unit ramp rate, and valve point effect, and to consider the requirement of minimizing the cost of multi-area power load comprehensively. Based on data-driven, this paper adopts an improved fruit fly optimization algorithm to quickly find the global optimal solution. The calculations are performed by IEEE6 simulation test system, and the results verify the feasibility of the proposed algorithm. The improved fruit fly optimization algorithm is compared and analyzed with other algorithms considering the quality of the obtained solutions. The results show the effectiveness and superiority of the proposed algorithm in solving multi-area economic dispatching problems in real power systems.

KEYWORDS

multi-area economic dispatching, multi-source data-driven, clean energy systems, new power system, constraint planning

1 Introduction

In response to global climate change, the structural characteristics of the energy system are gradual shift to efficiency, cleanliness and sustainability (Li et al., 2021). A high proportion of wind and solar energy promote the low-carbon development of multi-energy system economy. But there are many problems. On the one hand, it is transmitted by the original system factors such as force limitation, electric energy storage characteristics, and low effectiveness of demand-side response interactive management have highlighted the problem of increasing the curtailment rate of renewable energy in the system. On the other hand, renewable energy sources such as wind and solar contribute to themselves. There is extreme volatility and intermittency that cannot be achieved smoothly and stably Large-scale grid integration. The force is increasing,

the regulation capacity is insufficient, and the renewable energy is further affected consumption has an impact. Therefore, how to coordinate the scale development of renewable energy. The issue of flexible scheduling of multi-energy systems is crucial (Wang et al., 2021; Zongxiang et al., 2022).

In power system operation, economic dispatch is an important optimization problem. The goal of power systems economic dispatch is to minimize the total cost of generation while meeting the constraints of a single region. In contrast, the economic dispatch of a multi-regional power system typically involves dividing the generator set into several interconnected generating zones. The dispatch model calculates the system's power generation capacity and the amount of electricity exchanged between regions, while meeting constraints such as power demand and motor characteristics, thereby minimizing the overall power generation capacity cost. The economic dispatch of the power system is an important link affecting the economic operation of the power system. This solution model reduces the power generation and operating costs of the system by rationalizing the output of each unit in the power system and optimizing the system operation and disassembly schedule (Ma et al., 2018; Ma et al., 2019; Zhang et al., 2022).

Currently, there are more algorithms on economic dispatch of power systems. Wang et al. (2022) proposed an improved state transition algorithm (MTSTA) based on mirror transformation and dynamic axes transformation, and a new repair method of constraint handling combined with penalty function was used to deal with constraint conditions. Barukčić et al. (2022) adopted NSGA-II algorithm and considered the wind power fluctuations on the dynamic economic dispatch of spinning reserve constraints. As a commonly used optimization algorithm for solving complex problems, the swarm intelligence algorithm has greater advantages in the optimization of economic dispatch of power systems. The swarm intelligence algorithm solves the optimal solution for the economic dispatch of the power system by simulating individual generating units as bionic individuals and combining the power consumption path and demand. Xiao-hong and He (2017) proposed gravitational search-particle swarm optimization algorithm, and the individual with the largest particle fitness was obtained. According to the optimal individual position (Chen et al., 2022), the optimal economic dispatching scheme of power systems was obtained. Aiming at the problems of premature convergence in the traditional particle swarm optimization algorithm, a multi-agent particle swarm optimization algorithm based on chaos is used to solve it. Thus, the paper established a dynamic optimization dispatching model for power system with system frequency regulation constraints (Zhang and Ma, 2023). Author J. Yang, J. Liu, Y. Xiang, S. Zhang and J. Liu proposes a real-time dynamic scheduling strategy considering economic operation and complementary regulatory capabilities. They considered the uncertainty of photovoltaic and load demand, and studied the integrated power system of zero-carbon hydropower station (PV) pumped storage (PHS). The power fluctuation of the upper grid co-coupling point (PCC) after PHS participation is alleviated (Yang et al., 2023). In Literature (Pradeep Kumar and Pillai, 2020), V. Pradeep Kumar and A. S. Pillai use dynamic schedulers to compare static scheduling. Discover that dynamic scheduling adds flexibility and time-constrained guarantees. The author's study provides a comparison of the performance of fixed-priority and dynamic priority scheduling algorithms for automotive subsystems.

The economic dispatch problem is characterized by high-dimensional, non-convex, discrete, multi-constrained, and numerous local minima, which leads to a great difficulty in finding the optimal solution. The traditional bionic swarm intelligence algorithm is not prominent enough in local and global search capability, and the optimal extreme value solution is not accurate enough. In this paper, three test systems with different characteristics, IEEE6, IEEE40, and IEEE10, are selected and optimally scheduled using an improved fruit fly optimization algorithm. Among them, the IEEE6 machine test system and the IEEE40 machine test system are static scheduling models, and the IEEE10 machine test system is a dynamic scheduling model. Different test systems have different objective functions and constraints, and the treatment of the constraints in the specific optimization process will also be different.

2 Introduction to traditional economic scheduling

2.1 Objective function

The model has the optimization objective of minimizing the operating cost of the system, and its functional expression is Chen et al. (2022)

$$\min F_{\text{cost}} = \sum_{t=1}^T \sum_{i=1}^M F_i(P_{ti}) \quad (1)$$

Where, F_{cost} is the total system generation cost; T is the total number of dispatch periods. In dealing with static optimization problems, taking T to 1 is sufficient. M is the number of units in the system; P_{ti} is the active output value of unit i in period j ; $F_i(P_{ti})$ denotes the generation cost of unit i in time j .

In general, the generation cost of a thermal power unit can be expressed by its consumption characteristic function with the mathematical expression

$$F_i(P_{ti}) = a_i P_{ti}^2 + b_i P_{ti} + c_i \quad (2)$$

In this equation, a_i , b_i , c_i is the consumption characteristic coefficient of unit i .

When optimizing the two test systems, IEEE10 and IEEE40, threshold effects need to be considered. The valve point effect refers to the wire drawing effect that occurs when the turbine of a thermal power unit is suddenly opened by the intake valve (Xiao-hong and He, 2017). When considering the valve point effect, the traditional consumption characteristic function cannot accurately represent the input-output relationship of the unit. The solution is to superimpose a sinusoidal function on top of the traditional consumption characteristic function to correctly represent the power generation cost of thermal power units. If the valve point effect is taken into account, the objective function can be expressed as

$$\min F_{\text{cost}} = \sum_{t=1}^T \sum_{i=1}^M \{F_i(P_{ti}) + |e_i \sin[f_i(P_{i\min} - P_{ti})]|\} \quad (3)$$

Where, e_i , f_i is the valve point effect factor of unit i ; $P_{i\min}$ is the lower limit of active output of unit i .

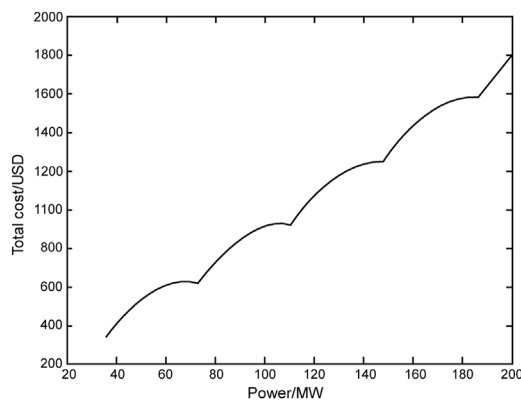


FIGURE 1
The total cost function of thermal power units taking into account the valve point effect.

The total operating cost curve of thermal power units considering the valve point effect is shown in Figure 1. It can be seen from the figure that due to the valve point effect, the originally smooth curve is non-convex. This leads to many local minima in the process of optimizing this test system, which increases the difficulty of optimization and puts high demands on the performance of the optimization algorithm.

2.2 Mathematical model of the traditional economic dispatching problem

2.2.1 System power balance constraint

The system power balance constraint is composed of the active output of the unit, the system network power loss and the total system load.

$$\sum_i^M P_{ti} - P_{tloss} - P_{tload} = 0 \quad (4)$$

Where P_{tloss} is the system network power loss at time t and P_{tload} is the total system load at time t . The system network power loss is obtained by the B-factor method as shown below.

$$P_{tloss} = \sum_{i=1}^M \sum_{j=1}^M P_{ti} B_{ij} P_{tj} + \sum_{i=1}^M B_{oi} P_{ti} + B_{oo} \quad (5)$$

The system network power loss is B_{ij} , B_{oi} , B_{oo} .

In practical simulations, since a strict system power balance is difficult to achieve, the general treatment is to set a very small value (Chen et al., 2022) ε ($0 \leq \varepsilon$). The power balance constraint is considered to be satisfied when the absolute value of the difference between the total output of the unit minus the network power loss and the load is less than ε . Meanwhile, the smaller the value of ε , the more difficult it is to satisfy the power balance constraint, and the more difficult it is to optimize.

2.2.2 Unit output constraint

$$P_{i\min} \leq P_i \leq P_{i\max} \quad (6)$$

$P_{i\min}$, $P_{i\max}$ are the lower limit of active output and the upper limit of active output of unit i , respectively.

2.2.3 Unit ramp rate constraints

Unit ramp rate constraints is an important constraint to be considered in dynamic dispatching. In dynamic dispatch, it is necessary to consider the upper and lower limits of unit output as well as the unit ramp rate constraints. The unit ramp rate constraints can be represented as (Zhang and Ma, 2023).

$$-DR_i \leq P_{ti} - P_{(t-1)i} \leq UR_i \quad (7)$$

Where: DR_i , UR_i are the power output growth rate extreme and power output decrease rate extreme of unit i respectively; $P_{(t-1)i}$ is the active output of unit i at time $t-1$.

2.2.4 Unit prohibited operating zones constraints

When a thermal power unit is in operation, there will be some subintervals within its operation interval. When thermal power units are operated within these subintervals, it will lead to excessive amplitude of unit bearing vibration. Therefore, it is necessary to set up prohibited operating zones within the operation interval to avoid these subintervals during the operation of the unit to prevent excessive vibration of the unit bearings. The operation interval with the prohibited operating zones set can be expressed as (Yang et al., 2023)

$$\begin{cases} P_{i\min} \leq P_i \leq P_{Si}^{d1} \\ P_{Si}^{h(j-1)} \leq P_i \leq P_{Si}^{dj} \\ P_{Si}^{hN_g} \leq P_i \leq P_{i\max} \end{cases} \quad (8)$$

Where, P_{Si}^{hj} is the lower limit of the j th prohibited operating zones of unit i ; P_{Si}^{dj} is the upper limit of the j th prohibited operating zones of unit i ; N_g is the total number of prohibited operating zones of unit i .

Due to the high parameter dimension and complex model state, this paper relies on experience to select parameters.

3 Constraint heuristic processing strategy

Most of the swarm intelligence optimization algorithms use the penalty function method when dealing with the constraints in constrained optimization problems (Hosseinneshad et al., 2014; Huo et al., 2015). The penalty function method requires an appropriate penalty factor to ensure the accuracy of the optimization and the efficiency of the whole optimization process when dealing with constraints. If the penalty factor is not set properly, it is very likely to make the optimization result less than expected and make the algorithm fall into local optimum in the process of finding the best. Therefore, this paper deals with the constraints through a heuristic processing strategy, thus avoiding unsatisfactory optimization results due to improper selection of penalty factors (Roy et al., 2014; Mishra et al., 2022). At the same time, different test systems consider different constraints and adjust the heuristic processing strategy as follows.

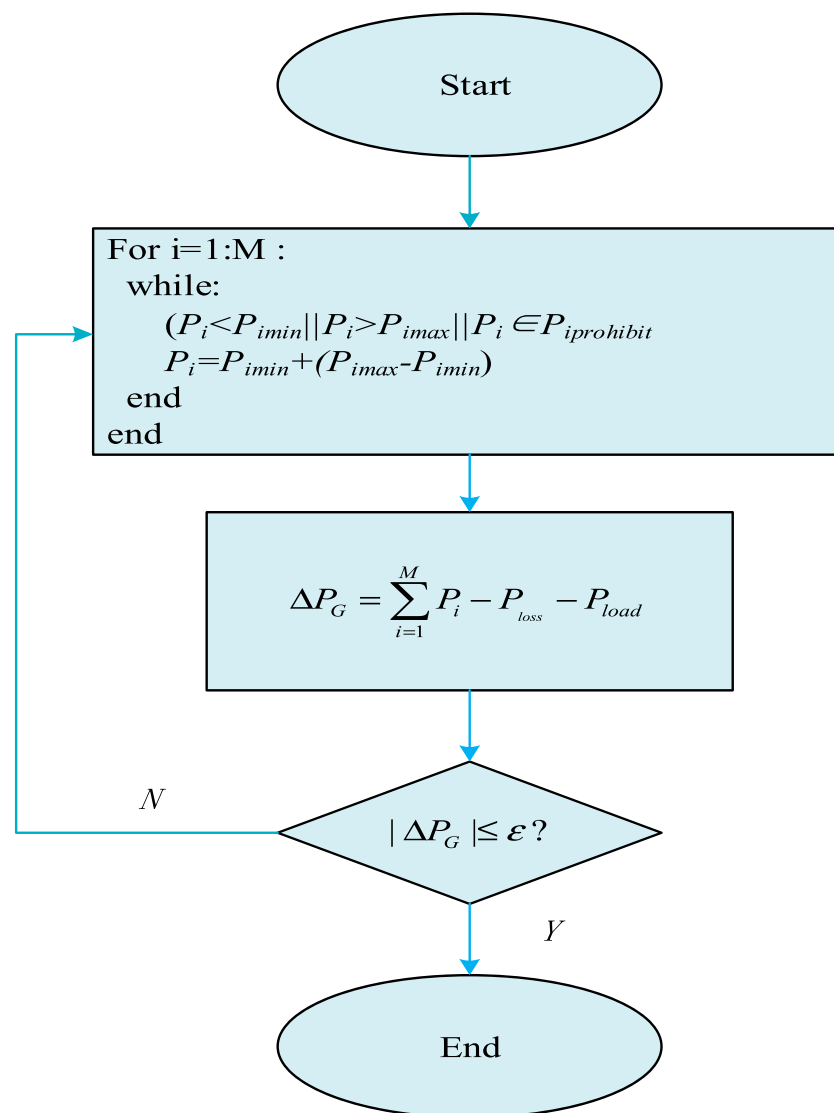


FIGURE 2
Heuristic processing flow chart for static dispatch constraints.

3.1 Strategies for handling static dispatching constraints

In this paper, we consider the system power balance constraint, unit output constraint, unit prohibited operating zones constraints and network power loss in optimizing the IEEE6 test system. In the optimization of the IEEE40 machine test system, the power balance constraint and the upper and lower limits of unit output are considered, and the network power loss and the operating exclusion zone constraint are ignored. The heuristic processing flow for the constraints of the two test systems is shown in [Figure 2](#). The specific steps are as follows ([Sa-Ngiamvibool et al., 2011](#)):

- Step 1: Determine whether any unit i satisfies the upper and lower limits of unit output and the prohibited operating zones constraints (When dealing with the constraints of the IEEE40 test system, only whether the upper and lower

output constraints are met is considered, and no prohibited operating zones constraints are considered.). The set of all prohibited operating zones of unit i is denoted by $P_i^{prohibit}$. If unit i cannot satisfy both the upper and lower unit output constraints and the prohibited operating zones constraints, the output value of unit i needs to be reset until both sets of constraints are satisfied.

- Step 2: Determine whether the system satisfies the power balance constraint. The inequality between unit output and network power loss and load of the test system is denoted by ΔP_G (In dealing with the power balance constraint of the IEEE40 test system, let $P_{loss} = 0$). To determine the degree of imbalance a minimal value $\varepsilon (0 \leq \varepsilon)$ needs to be set. When the absolute value of the imbalance is not greater than ε , the test system is considered to satisfy the power balance constraint. If the test system satisfies the power balance constraint, the algorithm outputs the calculated results. If the test system

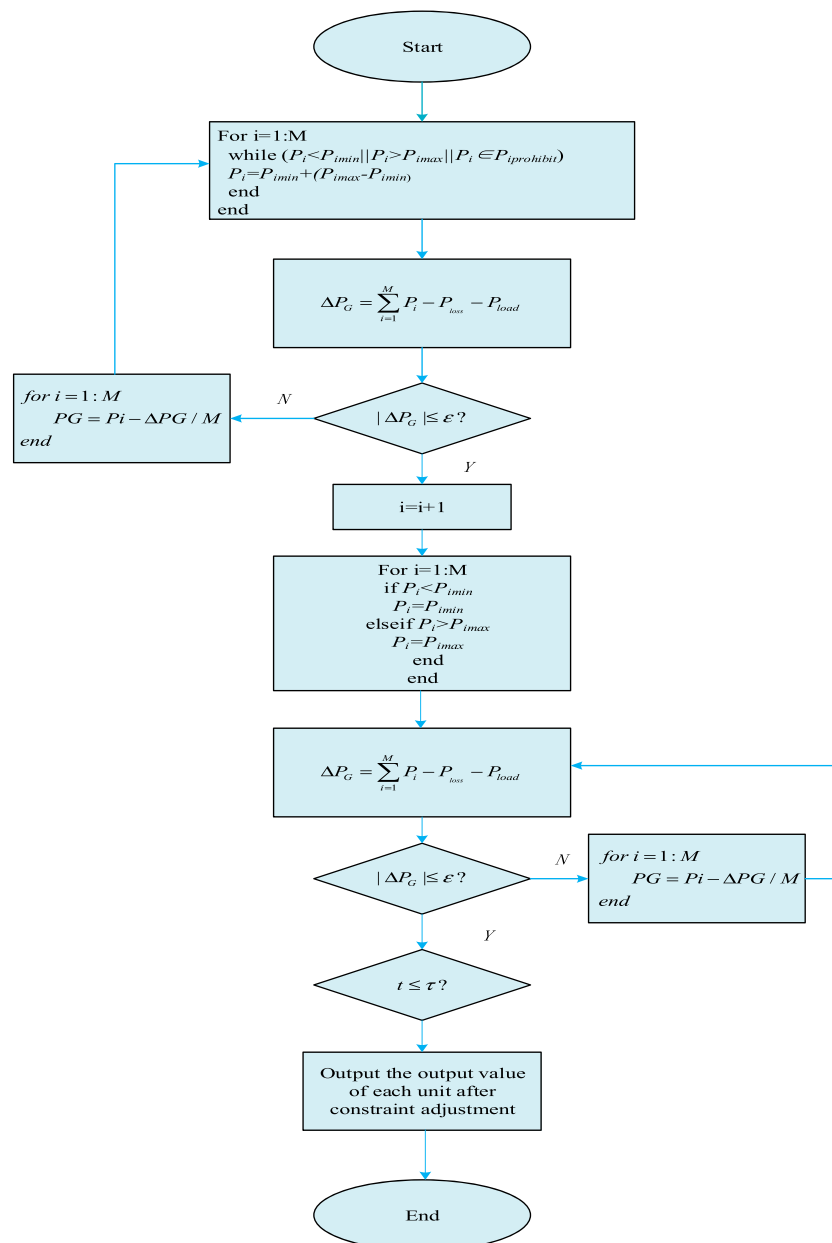


FIGURE 3
Heuristic processing flow chart for dynamic dispatch constraints.

does not meet the power balance constraint, the imbalance is distributed equally to each unit to adjust the unit output and network power loss, and then returns to Step 1.

3.2 Strategies for handling dynamic dispatching constraints

The system power balance constraint, unit output constraint, and unit climbing constraint are considered in optimizing the IEEE10 machine test system, and the network power loss is ignored. The heuristic processing flow of the constraints is shown in Figure 3. The specific steps are as follows:

- Step 1: Determine whether any unit meets the upper and lower capacity constraints at the first dispatch. If the constraint is not satisfied, the output value of the unit needs to be readjusted to meet the upper and lower limits of the unit output.
- Step 2: Determine whether the test system satisfies the power balance constraint at the first dispatch (set $P_{loss} = 0$) (Pradeep Kumar and Pillai, 2020). The algorithm uses P_{tG} to denote the system power imbalance at moment t . When the absolute value of the imbalance is not greater than ϵ , the test system is considered to satisfy the power balance constraint. If the test system does not satisfy the power balance constraint at the dispatching moment t , the imbalance is distributed equally to each unit to adjust the unit output and subsequently returns to

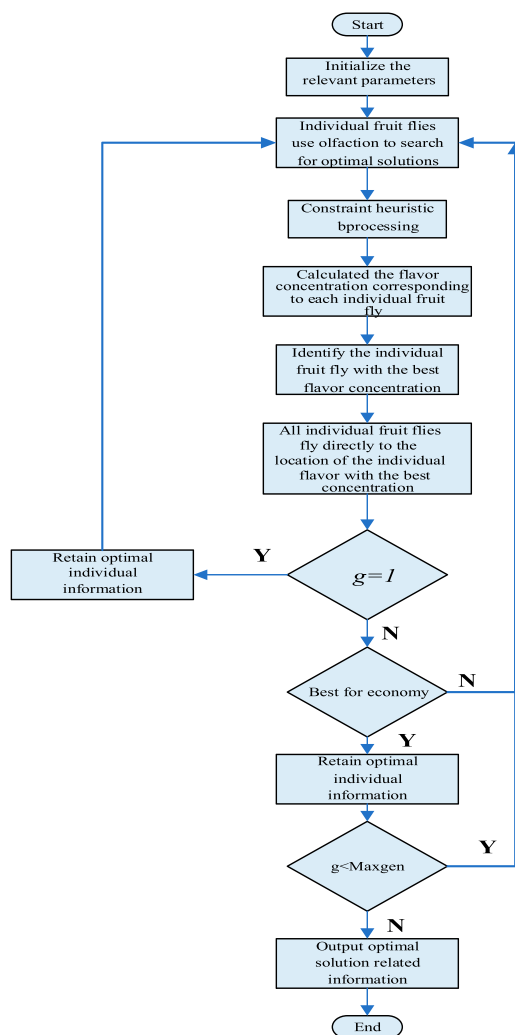


FIGURE 4
Flow chart of economic dispatch based on improved fruit fly optimization algorithm.

Step 1. If the test system satisfies the power balance constraint at the scheduling moment t , it goes to the next dispatching period.

- Step 3: Integrate the ramp rate constraints and the upper and lower unit output constraints into a set of constraints. The algorithm uses P_{iMIN} to denote the lower limit of unit output of unit i at dispatch time t and uses P_{iMAX} to denote the maximum unit output of unit i at dispatch time t after integration. Then go to the next step
- Step 4: Determine whether it is possible for unit i to satisfy the power balance constraint between P_{iMIN} and P_{iMAX} at the dispatch moment t . If this is not possible, the unit output values for the previous time period need to be adjusted to change P_{iMIN} and P_{iMAX} until it is possible for the test system to satisfy the power balance constraint for this time period within the constraints of P_{iMIN} and P_{iMAX} . If satisfied, go to the next step.
- Step 5: Determine whether any unit i satisfies the constraints in the range P_{iMIN} and P_{iMAX} at the dispatching time t . If the

constraint is not satisfied, the unit i needs to be readjusted output value at the dispatching moment t to satisfy the constraints in the range P_{iMIN} and P_{iMAX} . If the constraints are met, go to the next step.

- Step 6: Determine whether the test system satisfies the power balance constraint at the dispatching moment t . If the test system does not satisfy the power balance constraint at dispatch moment t , the imbalance is distributed equally to each unit to adjust the unit output, and then returns to Step 5. If the test system satisfies the power balance constraint at dispatch moment t , it proceeds to the next step.
- Step 7: Determine if the last dispatching moment has been reached. The algorithm uses T to represent the total number of scheduling periods for the test system. If the last dispatching moment has not been reached, return to Step 3 for the heuristic processing process of the constraints for the next dispatching moment. If the last dispatching moment has been reached, the unit output values for all dispatching moments are output.

4 Economic dispatch of power system based on improved fruit fly optimization algorithm

The algorithm flow is shown in Figure 4. The specific steps are shown in Figure 4.

- Step 1: Initialize the relevant parameters: set the total cost of system power generation F_{cost} as the fitness function. All units in the test system (the number of units is n) constitute one individual fruit fly, i.e., $X_i = \{P_i, \dots, P_n\}$. The algorithm sets the minimum search radius R_{min} to 1 and the maximum search radius R_{max} to the difference between the upper and lower limits of the output of the corresponding unit. The population size "Sizepop" is 50, and the maximum number of iterations "Maxgen" is set to 100 when optimizing the IEEE6 test system. When optimizing the IEEE40 test system, "Maxgen" is set to 300. When optimizing the IEEE10 test system, the total number of scheduling periods T is set to 24 and "Maxgen" is set to 300.
- Step 2: The individual fruit fly is given a search radius R as well as a random direction and distance to search using olfaction.
- Step 3: Execute a heuristic constraint processing strategy for each individual fruit fly so that each individual fruit fly satisfies the constraint.
- Step 4: Use the location $X_i = \{P_i, \dots, P_n\}$ of the individual fruit fly as the taste concentration determination value S_i . The algorithm brings the taste concentration determination value into the fitness function to find the taste concentration $Smell_i$ of individual fruit flies, which is the total system power generation cost.
- Step 5: Find the individual fruit fly with the best flavor concentration in the fruit fly population.
- Step 6: The taste concentration and location information of the optimal fruit fly individual are recorded and retained, and all fruit flies in the population use vision to fly to the location of the optimal fruit fly individual.

TABLE 1 Parameters related to the algorithm test system.

Unit number	Minimum power output/MW	Maximum power output/MW	a	b	c	Prohibited operating zones/MW
1	100	500	0.0070	7.0	240	[210 240], [350 380]
2	50	200	0.0095	10.0	200	[90 110], [140 160]
3	80	300	0.0090	8.5	220	[150 170], [210 240]
4	50	150	0.0090	11.0	200	[80 90], [110 120]
5	50	200	0.0080	10.5	220	[90 110], [140 150]
6	50	120	0.0075	12.0	190	[75 85], [100 105]

TABLE 2 Comparison of the optimal solutions of different algorithms.

Algorithms	Optimal solution/USD	Total output/MW	Network power loss/MW	ϵ /MW
Improved FOA	15442.661	1275.415	12.4149	0.0000
MABC Yu et al. (2022)	15449.8995	1275.958	12.9582	-0.0002
DE Lu et al. (2022)	15449.5826	1275.93	12.95	-0.02
KHA-IV Secui, (2015)	15443.0752	1275.445	12.4449	0.0001
GA-API Elsayed and El-Saadany. (2015)	15449.7	1275.97	12.98	-0.01
MIQCQP Mandal et al. (2014)	15443.07	1275.44	12.44	0.00
CSA Ciornei and Kyriakides. (2013)	15443.08	1275.447	12.447	0.000
λ -Consensus Mandal et al. (2014)	15452.09	1276.27	13.27	0.00
BBO Basu and Chowdhury. (2013)	15443.0963	1275.446	12.446	0.000
SA-PSO Basu and Chowdhury. (2013)	15447	1275.7	12.733	-0.033
HCRO-DE Binetti et al. (2014)	15443.0750	1275.4449	12.4449	0.0000

- Step 7: determines whether this iteration is the first one. If not, determine whether the best flavor concentration produced by this iteration is less than the best flavor concentration produced by the previous iteration (i.e., better economy). If the best flavor concentration produced by this iteration is greater than the best flavor concentration produced by the previous iteration, Step 2 is executed.
- Step 8: Iterate to find the best. If the current number of iterations is less than the maximum number of iterations “Maxgen”, Step 2 will be executed, if the current number of iterations is equal to the maximum number of iterations, the result will be output and the optimization process will be finished.

5 Example analysis

The example is based on the IEEE6 test system, and the system data are shown in Table 1. The total load of the system is 1260 MW, and each unit contains upper and lower output limits as well as two sets of prohibited operating zones constraints, taking into account the network power losses. The algorithm mentioned in this article does not guarantee convergence at any parameter setting. In this paper, the algorithm is executed in

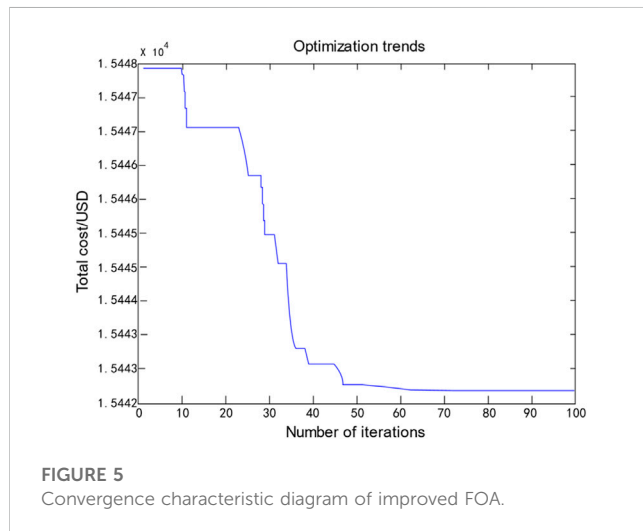
strict accordance with the scheduling optimization process, and the setting of each parameter is considered in detail to ensure that the model converges within a reasonable range to ensure that other performance of the model is not affected.

The unit contains upper and lower output limits and prohibited operating zones constraints, resulting in a discontinuous and non-convex solution space for this test system. $\epsilon(0 \leq \epsilon)$ reflects the required accuracy of the optimization result, the closer ϵ is to zero, the higher the accuracy of the optimization result. Table 2 shows the comparison of the optimal solutions of different algorithms. Equation 8 shows the output values of each unit after FOA (Fruit Fly Optimization Algorithm) optimization for different accuracy requirements (i.e., different values of A) in Table 3. Figure 5 shows the convergence characteristics of the improved FOA.

From Table 2, it can be observed that the optimal solutions of MIQCQP, CSA, λ -Consensus, BBO, and HCRO-DE are 15443.07USD, 15443.08USD, 15452.09USD, 15443.0963USD, and 15443.0750USD, respectively, with the same accuracy requirements, which are greater than the improved FOA's optimal solution of 15442.661USD. These algorithms, MABC, DE, KHA-IV, GA-API, SA-PSO, do not require as much accuracy as the improved FOA algorithm in the optimization process, but the optimal solution

TABLE 3 Optimization results with different solving accuracy.

ε	Total Output/MW	Network power loss/MW	Optimal solution/USD
0.0000	1275.415	12.4149	15442.661
0.01	1275.397	12.4065	15442.533
0.05	1275.352	12.4011	15442.021

**TABLE 4 P1-P6 Optimization results with different solving accuracy.**

ε	P1	P2	P3	P4	P5	P6
0.0000	447.1360	173.2631	263.9407	139.0616	165.4598	86.5537
0.01	447.2036	173.2650	264.2865	139.3577	164.9979	86.2859
0.05	446.0368	174.4044	263.4320	139.3787	165.1412	86.9587

obtained with the improved FOA is still the smallest. Also, as can be seen from Tables 3, 4, if the requirement for solution accuracy is not high when scheduling the optimization with the improved FOA, the resulting optimization results will be smaller. This shows that the improved FOA outperforms the other algorithms in terms of the solution quality of the optimal solution and the demanding degree of accuracy required.

As can be seen from Figure 5, the improved FOA shows good convergence in the search for the best FOA. In terms of optimization time, while optimizing the IEEE6 test system, the time required to improve FOA is 2.21 s, while the time required for SA-PSO and HCRO-DE is 7.58 s and 4.17 s, respectively. In terms of the number of iterations, the improved FOA only needs 73 iterations to reach the optimal value, while the CSA requires more than 100 iterations to reach the optimal value.

Taking into account the quality of the resulting solution, the improved fruit fly optimization algorithm is compared and analyzed with other algorithms. The results show that the algorithm has good effectiveness and superiority in solving the problem of multi-regional economic dispatch of actual power system.rior in solving the multi-regional economic dispatching problem of the actual power system.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

JS mainly establishes the implementation ideas of the paper and the detailed strategy of algorithm optimization in this paper, author CX models the paper and analyzes the results to reach the final conclusion, and the other authors supplement and improve the subtle aspects such as the format and graphing of the paper. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the State Grid Headquarters Science and Technology Project (No. 5400-202122573A-0-5-SF).

Acknowledgments

Thanks to the support of by the State Grid Headquarters Science and Technology Project (No. 5400-202122573A-0-5-SF). Thanks to all the authors of this article for their contributions in this article.

Conflict of interest

Authors CX and LM were employed by the company State Grid Hubei Electric Power Company Limited Economic Research Institute. Author LX was employed by the company State Grid Hubei Electric Power Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Barukčić, M., Varga, T., Benšić, T., and Jerković Štil, V. (2022). Optimal allocation of renewable energy sources and battery storage systems considering energy management system optimization based on fuzzy inference. *Energies* 15 (19), 6884. doi:10.3390/en15196884
- Basu, M., and Chowdhury, A. (2013). Cuckoo search algorithm for economic dispatch. *Energy* 60 (7), 99–108. doi:10.1016/j.energy.2013.07.011
- Binetti, G., Davoudi, A., Lewis, F. L., Naso, D., and Turchiano, B. (2014). Distributed consensus-based economic dispatch with transmission losses. *IEEE Trans. Power Syst.* 29 (4), 1711–1720. doi:10.1109/tpwrs.2014.2299436
- Chen, Z., Wang, Z., Li, X., and Zhang, C. (2022). Economic dispatching of power systems based on gravitational search-particle swarm optimization algorithm. *J. Univ. Jinan(Science Technol.* 36 (5), 603–608.
- Ciornei, I., and Kyriakides, E. (2013). A GA-API solution for the economic dispatch of generation in power system operation. *IEEE Trans. Power Syst.* 28 (1), 570. doi:10.1109/TPWRS.2011.2168833
- Elsayed, W. T., and El-Saadany, E. F. (2015). A fully decentralized approach for solving the economic dispatch problem. *IEEE Trans. Power Syst.* 30 (4), 2179–2189. doi:10.1109/tpwrs.2014.2360369
- Hosseinneshad, V., Rafiee, M., Ahmadian, M., and Ameli, M. T. (2014). Species-based quantum particle swarm optimization for economic load dispatch. *Int. J. Electr. Power & Energy Syst.* 63 (12), 311–322. doi:10.1016/j.ijepes.2014.05.066
- Huo, Y., Jiang, P., Zhu, Y., Feng, S., and Wu, X. (2015). Optimal real-time scheduling of wind integrated power system presented with storage and wind forecast uncertainties. *Energies* 8 (2), 1080–1100. doi:10.3390/en8021080
- Li, Z., Chen, S., Dong, W., Liu, P., Du, E., Ma, L., et al. (2021). Low carbon transition pathway of power sector under carbon emission constraints. *Proc. CSEE* 41 (12), 3987–4001. in Chinese.
- Lu, J., Xu, L., Jin, J., and Shao, Y. (2022). A mixed algorithm for integrated scheduling optimization in AS/RS and hybrid flowshop. *Energies* 15 (20), 7558. doi:10.3390/en15207558
- Ma, H., Wang, B., Gao, W., Zhu, C., Zhu, C. L., Zheng, Z., et al. (2019). Operation optimization of energy storage equipment participating in auxiliary service in regional integrated energy system. *Automation Electr. Power Syst.* 43 (8), 34–40+68.
- Ma, H., Wang, B., Gao, W., Liu, D., Sun, Y., and Liu, Z. (2018). Optimal scheduling of an regional integrated energy system with energy storage systems for service regulation. *Energies* 11, 195. doi:10.3390/en11010195
- Mandal, B., Roy, P. K., and Mandal, S. (2014). Economic load dispatch using krill herd algorithm. *Int. J. Electr. Power & Energy Syst.* 57 (5), 1–10. doi:10.1016/j.ijepes.2013.11.016
- Mishra, P., Miller, E., Santhanagopalan, S., Kevin, B., and Meintz, A. (2022). A framework to analyze the requirements of a multiport megawatt-level charging station for heavy-duty electric vehicles. *Energies* 15 (10), 3788. doi:10.3390/en15103788
- Pradeep Kumar, V., and Pillai, A. S. (2020). “Dynamic scheduling algorithm for Automotive safety critical systems,” in 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, March 11–13, 2020, 815–820.
- Roy, P. K., Bhui, S., and Paul, C. (2014). Solution of economic load dispatch using hybrid chemical reaction optimization approach. *Soft Comput.* 24, 109–125. doi:10.1016/j.asoc.2014.07.013
- Sa-Ngiamvibool, W., Pothiya, S., and Ngamroo, I. (2011). Multiple tabu search algorithm for economic dispatch problem considering valve-point effects. *J. Electr. Power & Energy Syst.* 33 (4), 846–854. doi:10.1016/j.ijepes.2010.11.011
- Secui, D. C. (2015). A new modified artificial bee colony algorithm for the economic dispatch problem. *Energy Convers. Manag.* 89 (89), 43–62. doi:10.1016/j.enconman.2014.09.034
- Wang, H., Yue, C., Yan, X., Hou, J., Liu, S., Zhao, T., et al. (2021). Influence of system inertia on flexibility resource analysis for an interconnection system with a high proportion of intermittent renewable energy. *Power Syst. Prot. Control* 49 (18), 44–51. in Chinese.
- Wang, P.-k., Wang, C., and Zhang, H.-L. (2022). Dynamic economic dispatching of power system with wind farms based on improved state transition algorithm. *Comput. Simul.* 39 (9), 130–135+155.
- Xiao-hong, H. A. O., and He, Kan (2017). Dynamic economic dispatch considering wind power penetration based on NSGA-II. *Electron. Des. Eng.* 25 (11), 170–175.
- Yang, J., Liu, J., Xiang, Y., Zhang, S., and Liu, J. (2023). Data-driven optimal dynamic dispatch for hydro-PV-PHS integrated power systems using deep reinforcement learning approach. *CSEE J. Power Energy Syst.* 9 (3), 846–858. doi:10.17775/CSEEJPES.2021.07210
- Yu, Q., Dong, Y., Du, Y., Yuan, J., and Fang, F. (2022). Optimizing operation strategy in a simulated high-proportion wind power wind-coal combined base load power generation system under multiple scenes. *Energies* 15 (21), 8004. doi:10.3390/en15218004
- Zhang, J., Wang, B., Wang, H., Ma, H., Ma, F., Li, Y., et al. (2022). Operation state evaluation method of Smart distribution network based on free probability theory. *Front. Energy Res.* 9, 803010. doi:10.3389/fenrg.2021.803010
- Zhang, L., and Ma, Y. (2023). Dynamic optimal dispatch of power system considering system frequency regulation constraints. *Electron. Des. Eng.* 31 (2), 188–193. doi:10.14022/j.issn1674-6236.2023.02.040
- Zongxiang, L. U., Yisha, L. I. N., and Ying, Q. (2022). supply-demand balance in power system with ultra-high proportion of renewable energy. *Automation Electr. Power Syst.* 46 (16), 3–16. in Chinese.



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Keyu Wu,
Institute for Infocomm Research (A*STAR),
Singapore
Liao Wu,
University of New South Wales, Australia

*CORRESPONDENCE

Hui Cao,
✉ huicao@mail.xjtu.edu.cn

RECEIVED 18 November 2023

ACCEPTED 21 December 2023

PUBLISHED 09 January 2024

CITATION

Li Q, Liu D, Cao H, Liao X, Lai X and Cui W (2024),
Decentralized asynchronous adaptive
federated learning algorithm for securely
prediction of distributed power data.
Front. Energy Res. 11:1340639.
doi: 10.3389/fenrg.2023.1340639

COPYRIGHT

© 2024 Li, Liu, Cao, Liao, Lai and Cui. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Decentralized asynchronous adaptive federated learning algorithm for securely prediction of distributed power data

Qiang Li¹, Di Liu¹, Hui Cao^{2*}, Xiao Liao¹, Xuanda Lai² and Wei Cui¹

¹State Grid Information & Telecommunication Group Co., Ltd., Beijing, China, ²School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China

Introduction: Improving the precision and real-time speed of electricity data prediction while safeguarding data privacy and security holds immense significance for all power system participants' decision-making. To surmount the issues of exorbitant computational expenses and privacy breaches of traditional centralized prediction methods, this paper proposes a decentralized asynchronous adaptive federated learning algorithm for securely prediction of distributed power data, which makes predictions from distributed data more flexible and secure.

Methods: First, each regional node trains its own deep neural network model locally. After that, the node model parameters are uploaded to the decentralized federated learning chain for ensuring local data protection. Asynchronous aggregated update of the global prediction model is then achieved via block mining and shared maintenance. The algorithm has been enhanced based on the traditional federated learning algorithm, which introduces an asynchronous mechanism while adaptively adjusting the regional node model weights and local update step size to overcomes the inefficiency of traditional methods.

Results and Discussion: The experimental analysis of actual electricity price data is conducted to compare and analyze with the centralized prediction model, study the impact of model adoption and parameter settings on the results, and compare with the prediction performance of other federated learning algorithms. The experimental results show that the method proposed in this paper is highly accurate, efficient, and safe.

KEYWORDS

time-series data prediction, distributed learning, federated learning, decentralization, privacy preservation

1 Introduction

Power time-series data has a significant impact on power system operation, planning and decision making. It can support decision making and optimization in load forecasting and dispatching, fault detection and handling, energy planning and market trading. However, with the development of distributed energy resources, the rise of microgrids and the Energy Internet, and the application of intelligent and digital technologies, several factors have contributed to making power systems become distributed (Liu et al., 2021). Meanwhile, power industries, which are a vital part of national energy security, have data on energy supply and demand, grid stability, etc. Therefore, it is necessary to develop an efficient method applicable to the analysis and prediction of distributed power data, which

can reduce the cost of data transmission and processing, and also ensure the security of private data in the power industry (Ali et al., 2023).

Currently, researchers have conducted relevant studies on the analysis and prediction of time series data in distributed power systems. The primary methods employed encompass traditional time series analysis and intelligent data mining techniques. In the time series approach, statistical analysis is conducted on power time series data to identify its characteristics, patterns, and trends, which are then utilized to derive predictive values (Badhiye et al., 2022; Frizzo et al., 2023). Some of the more classical time series models include Autoregressive (AR), Moving Average (MA), and Autoregressive Moving Average (ARMA). Furthermore, data mining methods are employed to capture potential non-linear relationships in the formation process of power time series data, thus enhancing prediction accuracy (Wang et al., 2023; Zhang et al., 2023). Commonly utilized intelligent methods include Random Forest (RF), Support Vector Machine (SVM), Deep Neural Network (DNN), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), among others.

The above methods for predicting electricity time series data usually adopt a centralized computing model, collecting all the data and concentrating them in a central node for model training and prediction analysis. During the training of the centralized predictive model, all data is transmitted and stored on the central server, the model is trained on the central server, and the update of the model is applied directly on the central server (Ahmed et al., 2022). However, the transmission of large amounts of data during centralized learning poses a risk to user privacy (Mcmahan et al., 2017). Meanwhile, the centralized mode has some problems, such as high data acquisition cost, low real-time performance, large resource consumption, and insufficient adaptability to the rapid change of the distributed power system (Li et al., 2020a).

To overcome the shortcomings of centralized tariff prediction models, some studies have applied techniques such as distributed computing and federated learning to power systems, which delegate data processing and analysis tasks to multiple edge nodes to improve efficiency, real-time performance and security. Federated learning methods can perform model training and parameter updating without exposing data and achieve better privacy results. For example, in FedAvg (Mcmahan et al., 2017), participating nodes perform local training and upload model parameters to the server, which performs parameter aggregation and model updating and distributes the global model to nodes, iteratively performing the above steps until convergence.

However, traditional federated learning methods do not perform well in the presence of heterogeneity in the system (Li et al., 2020b). In a fluctuating and undulating distributed power system, the conditions in each region are heterogeneous, and there may be problems such as inconsistent data distribution in each region, different data volumes, different data transmission efficiencies, and different computing power (Yu et al., 2019; Zeng et al., 2023). If an inappropriate data analysis and prediction method is used, the results obtained will have a negative impact on strategy analysis and decision making.

To address the limitations of the above methods, this paper proposes a decentralized asynchronous adaptive federated learning

algorithm (DAAFed) for securely prediction of distributed power data. The main contributions of this study are summarized as follows.

- 1) A decentralized distributed framework has been developed for predicting regional time series data. Each regional node trains its own local data prediction model using its unique data. Next, the parameters of the deep network model are uploaded. With the use of blockchain technology, the aggregation update of the global prediction model is realized through the generation of blocks and the common maintenance of block sets.
- 2) The conventional federated learning algorithm is improved by the integration of a time-synchronous mechanism, which adopts the adaptive adjustment of the regional node model weight and the local update step size, thereby improving the accuracy and efficiency of distributed data prediction.
- 3) The proposed method will be validated using measured data from different regions in Spain. By comparing and analyzing the results with a centralized prediction model, investigating the impact of model adoption and parameter settings on the results, and comparing the prediction performance with other federated learning algorithms, the experimental results validate that the proposed method is able to balance safety, accuracy and efficiency.

The rest of this article is organized as follows. Sections 2, 3 present a review of related work and the basic model for power time series data prediction, respectively. The decentralized asynchronous adaptive federated learning method proposed in this article is explained in Section 4. Then, in Section 5, we perform simulation experiments to verify the effectiveness of the method proposed in this paper. Finally, Section 6 concludes this article.

2 Related work

In distributed machine learning, federated learning, as an emerging artificial intelligence technology, can ensure data privacy while performing efficient machine learning, providing a new way to solve the “data silo” problem. Federated learning algorithms have been gradually applied to various problems, such as healthcare (Chen et al., 2023), communication (Qu et al., 2023), language modelling (Wu et al., 2020), transportation (Qi et al., 2023), etc. Federated learning was proposed by Google in 2017, where global models are trained through the cooperation of edge devices without sharing training data. In this approach, training is performed by edge devices, and the weights of the training results are shared with a central server to perform weight updates. And the updated weights are then sent back to the edge devices for a new round of training.

The basic federated learning methods are only suitable for certain environments, so there are also some improved federated learning methods that can be useful in different environments. A federated learning approach in conjunction with deep autoencoder networks based on representation learning was proposed by Husnoo et al. (2023) to enable monitoring and data collection subsystems in distributed grid regions to collaboratively train attack detection models for accurate detection of power system and network

security anomalies without sharing sensitive power-related data. The K-means was used to cluster power data locally at the utility, followed by federated learning to build accurate power prediction models for each class of power data in conjunction with other local clients (He and Zhao, 2022). A combined federated learning and deep reinforcement learning scheme for ultra-short term wind power prediction has been proposed by Li et al. (2023a), which uses a deep deterministic policy gradient algorithm as the basic prediction model to improve prediction accuracy, and then integrates the prediction model into the federated learning framework.

Recently, blockchain-based federated learning approaches have become increasingly popular. Blockchain provides tamper-proof and non-repudiation guarantees for the data stored in its ledger. Therefore, storing local and global models in the blockchain helps to improve the overall integrity of the system. Blockchain nodes can also collaborate to maintain the stored data and detect any malicious intentions of the participants. In addition, the execution of smart contracts is deterministic, which allows the federated learning process to be executed correctly and fairly. Intelligent encryption was redesigned to take advantage of the decentralized nature of blockchain technology to design collaborative intelligent federated learning frameworks for automated diagnosis without violating the trustworthiness metrics of privacy, security and data sharing encountered in smart city healthcare systems (Mohamed et al., 2023). Privacy-sensitive energy data can be stored locally at edge producer-consumer nodes without disclosure to third parties, and only the learned local model weights are shared using a blockchain network. Smart contracts are also used to handle the integration of local machine learning predictive models with the blockchain, enabling model parameter scaling and reducing blockchain overhead (Antal et al., 2022). A blockchain federated learning based object detection scheme was proposed by Li et al. (2023b), which eliminates central authority by using a distributed InterPlanetary File System (IPFS). The global model is periodically aggregated when multiple local model parameters are uploaded to the IPFS. Nodes can retrieve the global model from the IPFS. This method has been used in face detection, animal detection, unsafe content detection, vehicle detection, etc.

3 Preliminary

The characteristics and patterns of electricity time series data are intricately linked to the geographical location and regional environment of each distributed area. Factors such as electricity demand, load peaks and troughs, weather conditions, power generation structure, and network configuration in different regions all contribute to the distinct characteristics of electricity data. To address the challenge of predicting electricity time series data in diverse regions and to more effectively capture the non-linear relationships and data features of electricity prices, this study adopts a deep neural network model for accurate predictions at the edge nodes. Specifically, we focus on investigating and analyzing the commonly used CNN and LSTM, as well as their combination, to achieve accurate forecasting.

Let Ω_k represent the model of the k th node, and the sample data used for training the prediction includes various feature data such as load information L_k , generation information G_k , weather information

W_k , date information D_k , along with the corresponding labeled data for prediction target P_k . The representation can be expressed as Eqs 1, 2:

$$\mathbf{x}_k = [L_k, G_k, W_k, D_k] \quad (1)$$

$$\mathbf{y}_k = [P_k] \quad (2)$$

Training of each node is based on both the model as well as the sample data. The prediction accuracy is higher when the value of the loss function is small. The commonly used loss function, Root Mean Square Error (RMSE), can be expressed as in Eq. 3:

$$l(\Omega(\mathbf{x}_k), \mathbf{y}_k) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\Omega(\mathbf{x}_i) - \mathbf{y}_i)^2} \quad (3)$$

where n is the total number of samples at the k th node, $\mathbf{x}_i \in \mathbf{x}_k$, $\mathbf{y}_i \in \mathbf{y}_k$, and $\Omega(\mathbf{x}_i)$ is the prediction output of the model.

Considering the efficient utilization of computing resources in each region and the paramount importance of data security, this study adopts the federated learning framework for model training and parameter updating in the region-specific prediction. Each edge node possesses data transmission and computation capabilities, making it advantageous to delegate computation tasks to the edge nodes, thereby ensuring accurate and efficient time series data prediction in a trustworthy environment.

In contrast to the traditional centralized learning model that uploads raw data to a central server for model training, the federated learning framework assigns the model training task to distributed local devices. Subsequently, the global model is updated on the server side by exchanging certain parameters. The process of prediction for each region using the federated learning framework comprises the following steps:

- 1) The server defines the general task of prediction, including the determination of global variables such as the choice of prediction model, training rounds, and aggregation rounds.
- 2) As in Eq. 4, each node determines the initial local model based on the global model ω . The node then trains the local model Ω_k using local data, encompassing generation, load, climate, and date information. The node then uploads the parameters \mathbf{v}_k , such as local model weights, to the server.

$$\mathbf{v}_k \leftarrow \text{LocalUpdate}(\omega, \mathbf{x}_k, \mathbf{y}_k), k = 1, 2, \dots, K \quad (4)$$

- 3) As in Eq. 5, the server globally aggregates the model parameters from each node according to the specified mechanism to construct a new global model ω .

$$\omega \leftarrow \text{GlobalUpdate}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K) \quad (5)$$

- 4) Each node receives the latest global model as the initial model for the next round of training and iteratively performs steps 2) and 3) until the model prediction reaches the desired accuracy.

4 Methodology

In this section, a decentralized asynchronous adaptive federated learning algorithm is proposed, which is oriented towards the secure prediction of distributed power data. As shown in Figure 1, the

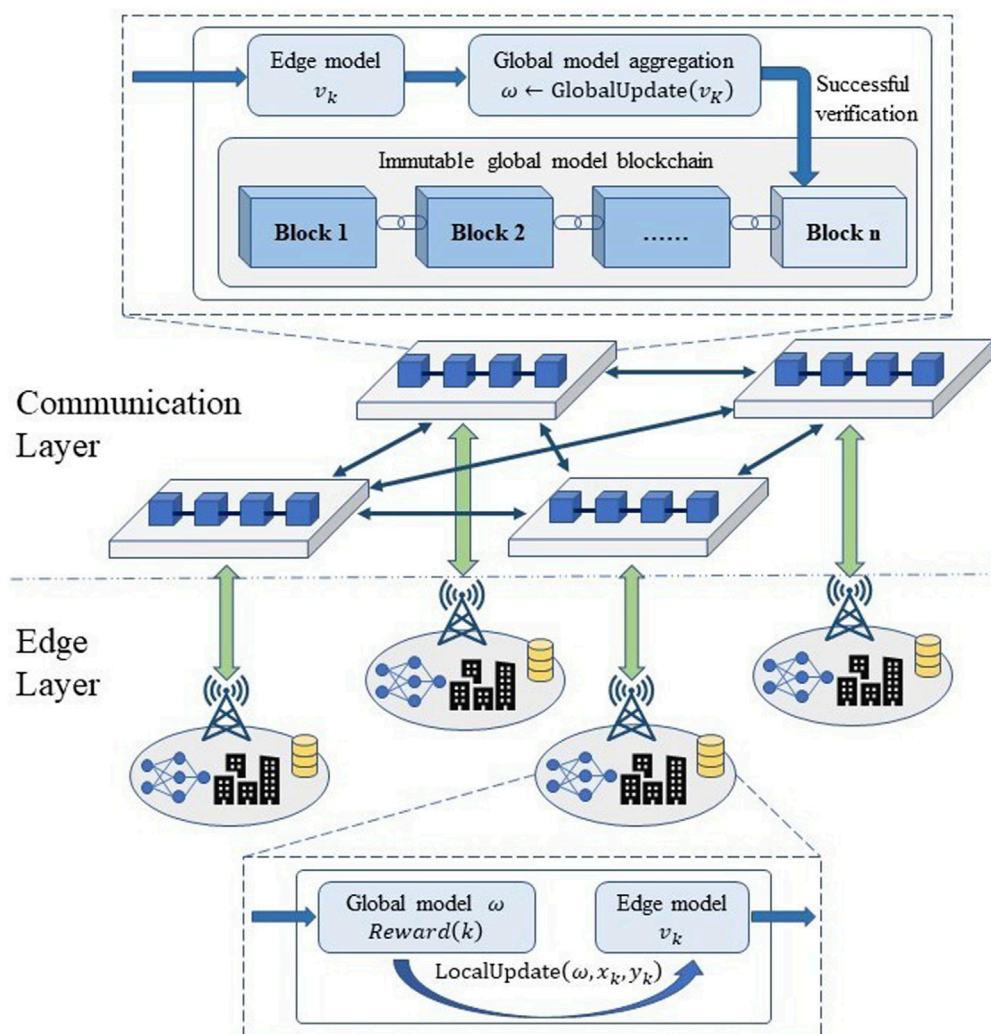


FIGURE 1
Distributed regional electricity price forecasting framework.

distributed regional time-series data prediction framework designed in this paper has a two-layer structure. The edge layer consists of each actual compute node, including the local power network and real-time data. The communication layer is the communication link between nodes, which is used to aggregate and update the model under the cooperation. First, each regional node locally trains its own deep network model Ω_k and uploads the node prediction model parameters v_k to the federated learning chain to realize local data protection. Then, the global model ω is updated asynchronously based on the distributed ledger mechanism in blockchain technology, and the aggregation of global models is realized by block mining, and each regional node obtains the latest global model ω through the shared immutable blockchain for subsequent prediction tasks.

4.1 Traditional federated learning algorithm

Within the federated learning framework, each participating node conducts local training and uploads model parameters to the server. The

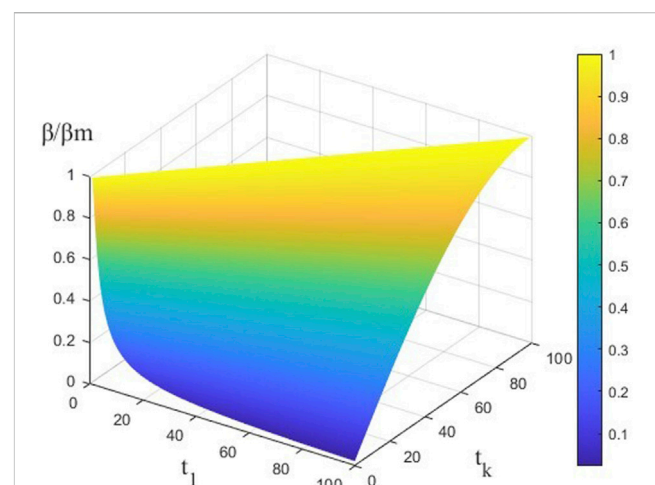


FIGURE 2
Weight adaptive adjustment function.

TABLE 1 Pseudocode of DAAGFed.

Algorithm	DAAGFed
Input:	K (number of the involved edge nodes) e (max number of iterations of the local model)
	E (max number of aggregations of the global model)
	S_k (data samples of each node)
Output:	v_k (local model parameter of each node)
	ω (global model parameter)
1	set global model parameters ω to initial values, set $v_k = \omega$
2	for $t = 0$: $(E - 1)$ do
3	# Parallel iterative phase at the edge nodes
4	parfor $k = 1$: K do
5	node#k updates the step-size η as in Eq. 28
6	node#k updates $v_k(t_1)$ and $\mu_k(t_1)$ as in Eqs 26, 27
7	node#k packages and broadcasts TX_0 as in Eq. 40
8	if the latest TX_0 is collected then
9	determine the corresponding β_k as in Eq. 20
10	perform global model aggregation $\omega(t_1)$ as in Eq. 19
11	package the global aggregation packet TX_1 as in Eq. 41
12	generate $Block_N$ and broadcast as in Eqs 42–46
13	end if
14	if a new block is received and verified then
15	add the block to the global model block set
16	update locally stored global model ω
17	update local task start moment t_k
18	go to step#5
19	end if
20	if no information feedback is collected then
21	go to step#5
22	end if
23	end parfor
24	end for

server aggregates parameters and updates the model, distributing the global model to each node. These procedures are iterative and continue until convergence is achieved. The most commonly used model aggregation algorithm for federated learning is FedAvg. The total number of nodes is K , the k th node contributes data samples as S_k , then the number of samples of this node is $|S_k|$, and the total number of data samples is $S = \sum_{k=1}^K |S_k|$.

As in Eq. 6, the global objective function that requires optimization is $\min F(\omega)$, where $\omega \in R^d$:

$$F(\omega) = \frac{1}{S} \sum_{i=1}^S f_i(\omega) \quad (6)$$

where $f_i(\omega) = l(x_i, y_i; \omega)$ represents the prediction loss for the sample (x_i, y_i) under the model parameter ω , then Eq. 6 can be rewritten as Eqs 7, 8:

$$F_k(v_k) = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_i(v_k) \quad (7)$$

$$F(\omega) = \sum_{k=1}^K \frac{|S_k|}{S} F_k(v_k) \quad (8)$$

where $F_k(v_k)$ is the loss function of the k th node, while $F(\omega)$ is the overall federated loss function.

Each iteration process of FedAvg comprises a local update and a global update. During every iteration, the following process is undertaken:

Initially, the latest global model from the server is received as the local model for the current round during the local update of each node. Then, e iterations of the local model are carried out, updating the model parameter ω^k for the k th node in each iteration as in Eq. 9:

$$v_k^t = v_k^{t-1} - \eta \nabla F_k(v_k^{t-1}) \quad (9)$$

where, η is the learning rate of the local update at each node.

Secondly, once the local update is executed, the model parameters are uploaded to the server. Subsequently, a server-side global aggregation update is implemented using the following aggregation formula as in Eq. 10:

$$\omega_{t+1} = \sum_{k=1}^K \frac{|S_k|}{S} v_k^t \quad (10)$$

4.2 Asynchronous adaptive mechanism

The asynchronous adaptive federated learning algorithm proposed in this paper makes improvements based on the FedAvg algorithm. In the process of global model aggregation, an asynchronous mechanism and adaptive adjustment of node model weights based on the determination of time obsolescence are introduced. In the process of updating local node models, personalized learning and adaptive step length adjustment are introduced to ensure asynchronous real-time and regional personalization, and to improve the prediction accuracy while enhancing the generalization ability of the global model. The specific implementation process of the asynchronous adaptive federal learning algorithm is shown as follow.

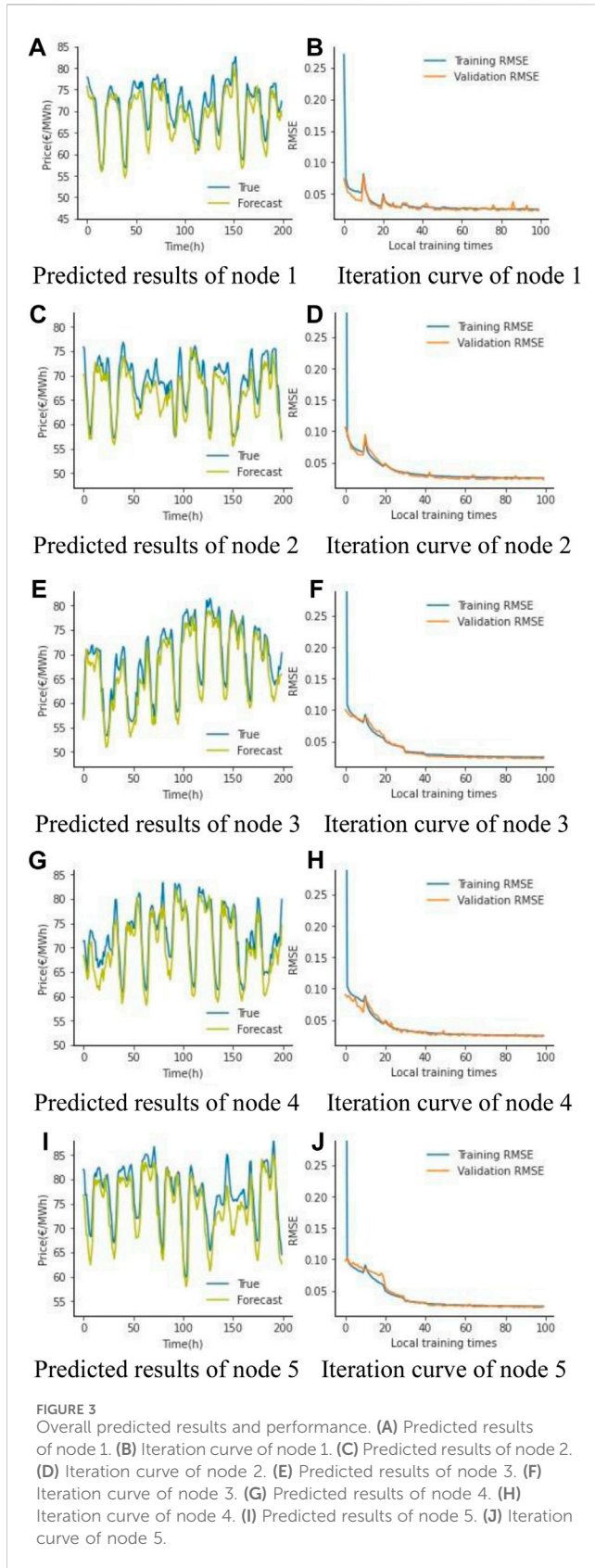
First, the overall objective function is determined based on Eq. 10. In the asynchronous adaptive federal learning algorithm, the overall objective function is Eq. 11:

$$\min \sum_{k=1}^K \frac{|S_k|}{S} F_k(v_k) \quad (11)$$

Let $\lambda_k = \frac{|S_k|}{S}$ denote the weight of each node. Then add model personalization constraints as in Eq. 12:

$$s.t. \|v_k - \omega\|^2 \leq T_k \quad (12)$$

where $\{T_k, k = 1, 2, \dots, K\}$ is the model personalization tolerance threshold, which is used to reflect the difference between the global model and the local model to achieve the personalized learning of the node's local model. The above optimization objective function with constraints can be rewritten by using the Lagrange equation as in Eq. 13:



$$L(\mathbf{v}_k, \boldsymbol{\omega}, \boldsymbol{\mu}_k) = \sum_{k=1}^K \lambda_k F_k(\mathbf{v}_k) + \frac{\boldsymbol{\mu}_k}{2} (\|\mathbf{v}_k - \boldsymbol{\omega}\|^2 - T_k) \quad (13)$$

The partial derivative of this objective function can be obtained:

$$\nabla_{\mathbf{v}_k} L = \lambda_k \nabla F_k(\mathbf{v}_k) + \boldsymbol{\mu}_k (\mathbf{v}_k - \boldsymbol{\omega}) \quad (14)$$

$$\nabla_{\boldsymbol{\omega}} L = \sum_{k=1}^K \boldsymbol{\mu}_k (\mathbf{v}_k - \boldsymbol{\omega}) \quad (15)$$

$$\nabla_{\boldsymbol{\mu}_k} L = \frac{1}{2} (\|\mathbf{v}_k - \boldsymbol{\omega}\|^2 - T_k) \quad (16)$$

where $k = 1, 2, \dots, K$.

Secondly, in the aggregation of global models, the asynchronous characteristics of multi-node collaboration should be considered. Asynchronous training allows each node to update model parameters independently with the server without worrying about the calculation pace of other nodes. In the case of asynchronous training, the model parameter update on the server side is carried out asynchronously, that is, there is no need to wait for the progress of other nodes that are being calculated. Considering the different time required for each node to complete the local task, this paper assumes that global aggregation begins whenever a single edge node completes uploading model parameters. Therefore, the variable $\varphi_k(t) \in \{0, 1\}$ is introduced to indicate the completion of node updates as in Eq. 17:

$$\varphi_k(t) = \begin{cases} 1, & \text{kth node completes update at } t \\ 0, & \text{others} \end{cases} \quad (17)$$

The above variables satisfy the constraints as in Eq. 18:

$$\sum_{k=1}^K \varphi_k(t) = 1 \quad (18)$$

According to Eq. 15, the global model aggregation is given by:

$$\boldsymbol{\omega}(t_1) = \boldsymbol{\omega}(t_0) + \sum_{k=1}^K \varphi_k(t_1) \beta_k(t_k, t_1) (\mathbf{v}_k(t_1) - \boldsymbol{\omega}(t_0)) \quad (19)$$

where t_1 is the generation moment of the global model of the current generation, which is also the moment when the local model of the k th node completes this update, t_0 is the generation moment of the global model of the previous generation, t_k is the generation moment of the global model adopted by the local model of the k th node when it performs the current update, β_k is the weight adaptive adjustment function based on the determination of the temporal staleness, which is determined according to the before and after of the local update of the k th node. If this model is outdated, the current node model is given a smaller weight, the specific form can be defined as:

$$\beta_k(t_k, t_1) = \beta_m \cos\left(\frac{\pi(t_1 - t_k)}{2(t_1 + \varepsilon)}\right) \quad (20)$$

where β_m is the maximum learning weight, ε is a constant value close to zero. And when t_k is similar to t_1 , β_k takes a value close to β_m .

$$\frac{\partial \beta_k}{\partial t_k} = \beta_m \sin\left(\frac{\pi(t_1 - t_k)}{2(t_1 + \varepsilon)}\right) \left(\frac{\pi}{2(t_1 + \varepsilon)}\right) > 0 \quad (21)$$

$$\frac{\partial \beta_k}{\partial t_1} = -\beta_m \sin\left(\frac{\pi(t_1 - t_k)}{2(t_1 + \varepsilon)}\right) \left(\frac{\pi(t_k + \varepsilon)}{2(t_1 + \varepsilon)^2}\right) < 0 \quad (22)$$

$$\frac{\partial^2 \beta_k}{\partial t_k^2} = -\beta_m \cos\left(\frac{\pi(t_1 - t_k)}{2(t_1 + \varepsilon)}\right) \left(\frac{\pi}{2(t_1 + \varepsilon)}\right)^2 < 0 \quad (23)$$

$$\begin{aligned} \frac{\partial^2 \beta_k}{\partial t_1^2} = & -\beta_m \cos\left(\frac{\pi(t_1 - t_k)}{2(t_1 + \varepsilon)}\right) \left(\frac{\pi(t_k + \varepsilon)}{2(t_1 + \varepsilon)^2}\right)^2 + 2\beta_m \sin\left(\frac{\pi(t_1 - t_k)}{2(t_1 + \varepsilon)}\right) \\ & \times \left(\frac{\pi(t_k + \varepsilon)}{2(t_1 + \varepsilon)^3}\right) \end{aligned} \quad (24)$$

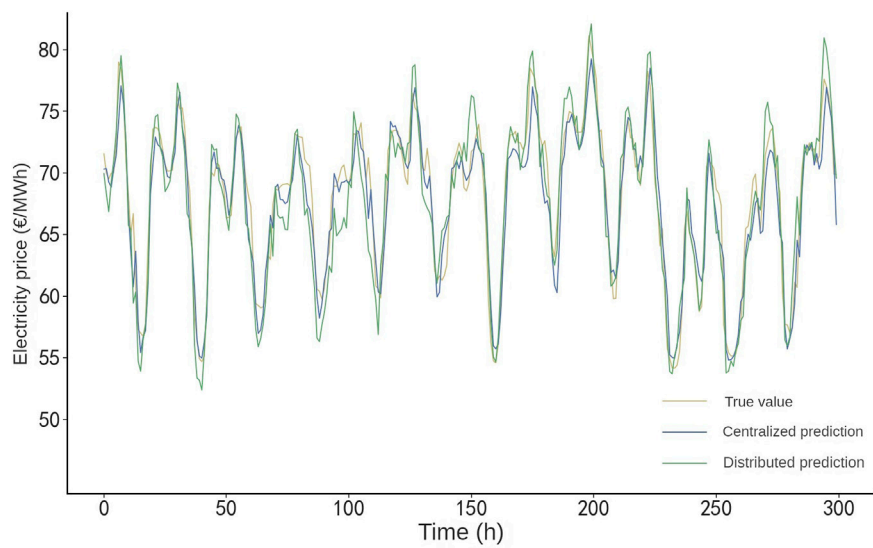


FIGURE 4
Comparison of centralized and distributed prediction.

TABLE 2 Comparison of centralized and distributed prediction.

Mode	Running time of each node (S)	RMSE	Data transmitted in a single transmission
Centralized	961.62	2.3	$5 \times 35,064 \times 33$
Distributed	817.79	2.313	53,601

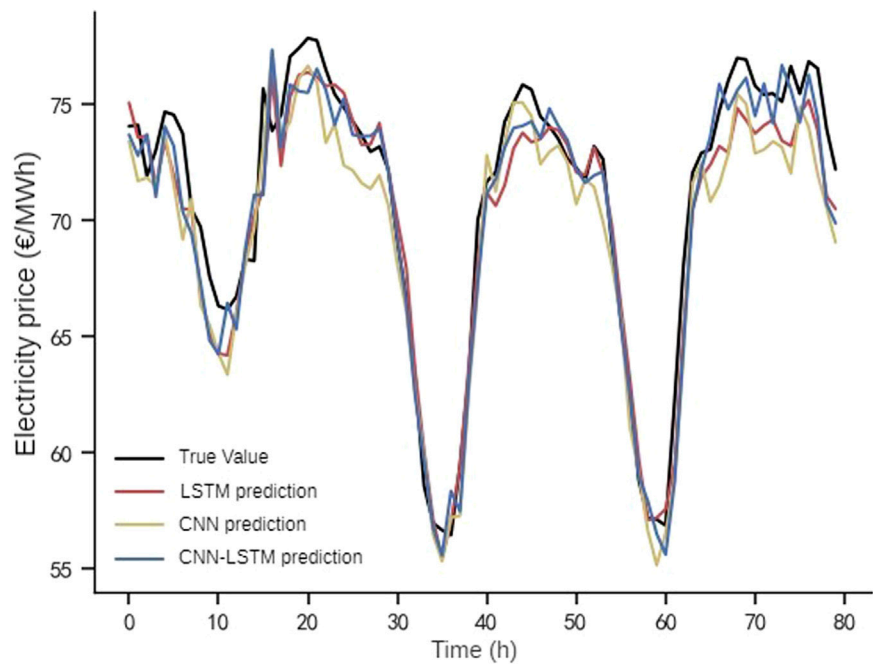


FIGURE 5
Comparison of prediction results under different models.

TABLE 3 Comparison of RMSE under different models.

Evaluation indicators	LSTM	CNN	CNN-LSTM
RMSE	2.377	2.353	2.313

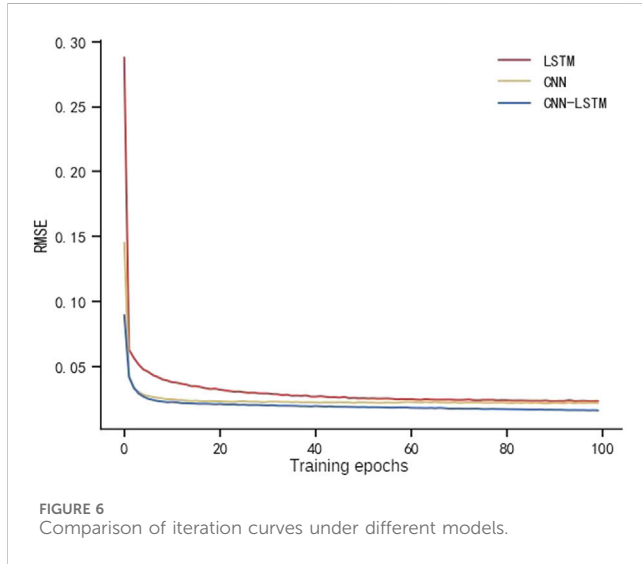


FIGURE 6 Comparison of iteration curves under different models.

The values of β_k vary with t_k and t_1 as shown in Figure 2. According to Eqs 21–24, β_k increases with the increase of t_k , and the magnitude of the increase decreases gradually, indicating that when the local model is updated with the newer global model, the weight of the updated node model should be increased accordingly. Also β_k decreases with the increase of t_1 . When t_k is smaller, $\frac{\partial^2 \beta_k}{\partial t_1^2} > 0$, which means β_k is convex. When t_k is larger, $\frac{\partial^2 \beta_k}{\partial t_1^2} < 0$, which means β_k is concave. It indicates that the global model parameters obtained at the early stage of prediction have large weights only at the early stage, and their weights decrease rapidly in the later stage of prediction.

Convergence analysis is carried out on the aggregation of the global model. When $T \rightarrow \infty$, the algorithm is considered convergent if the statistic $R(T) = \sum_{t=1}^T [F(\omega(t)) - F(\omega^*)]$ meets the predetermined convergence criteria that $R(T)/T \rightarrow 0$, which means $\omega \rightarrow \arg \min \sum_{t=1}^T F(\omega(t)) \triangleq \omega^*$, ω converges to ω^* to minimize the objective function.

The convergence analysis is shown as in Eq. 25.

$$\begin{aligned} R(T) &= \sum_{t=1}^T [F(\omega(t)) - F(\omega^*)] \leq \sum_{t=1}^T \langle \omega(t) - v, \omega(t) - \omega^* \rangle \\ &\leq \sum_{t=1}^T \frac{1}{2\beta_t} [\|\omega(t) - \omega^*\|^2 - \|\omega(t') - \omega^*\|^2] + \frac{\beta_t}{2} \|\omega(t) - v\|^2 \end{aligned} \quad (25a)$$

According to (22), it can be inferred that β_t decreases monotonically. Using the assumption that $\|\omega(t) - \omega^*\|^2 \leq D^2$ and local node personalization constraints $\|\omega(t) - v\|^2 \leq G^2$, the above equation can be simplified and transformed as follows:

$$\begin{aligned} R(T) &\leq \frac{1}{2\beta_1} D^2 + D^2 \sum_{t=2}^T \left(\frac{1}{2\beta_t} - \frac{1}{2\beta_{t-1}} \right) + \frac{G^2}{2} \beta_T \\ &= \frac{D^2}{2\beta_T} + \frac{G^2}{2} \beta_T \leq \frac{D^2}{2\beta_{\min}} + \frac{G^2 \beta_{\max}}{2} \end{aligned} \quad (25b)$$

Consequently, $R(T)$ has an upper bound and mean value of the statistic $R(T)/T \rightarrow 0$ when $T \rightarrow \infty$, ensuring the convergence of the global model aggregation.

Furthermore, for the update process of the edge model, according to Eqs 14, 16, the update process can be expressed as follow:

$$\begin{aligned} v_k(t_1) &= v_k(t_k) \\ &\quad - \eta_1 (\nabla F_k(v_k(t_k))) \times [\lambda_k \nabla F_k(v_k(t_k)) + u_k(v_k(t_k) - \omega(t_k))] \end{aligned} \quad (26)$$

$$u_k(t_1) = u_k(t_k) - \eta_2 (\|v_k(t_k) - \omega(t_k)\|^2) \times (\|v_k(t_k) - \omega(t_k)\|^2 - T_k) \quad (27)$$

where η_1 and η_2 are the adaptive adjustment function of the step size, which can be expressed as

$$\eta(z) = \max\{\eta_{\min}, \min\{\eta_{\max}, p^{qz}\}\} \quad (28)$$

where η_{\min} is the minimum value of the step, η_{\max} is the maximum value of the step, $p > 1$ and $q > 0$, z is the independent variable of the adaptive function, and $\eta(z)$ is a non-decreasing function taking values in the interval $[\eta_{\min}, \eta_{\max}]$. As in Eq. 28, it is indicated that when the gradient of the loss function is small and the local model is close to the global model, the step size is adaptively reduced to prevent non-convergence in the learning of the local model. Compared with the fixed step size, the adaptive step size can adjust the amplitude of the parameter changing direction along the gradient according to the training stage, and improve the convergence rate while ensuring the convergence accuracy.

Convergence analysis of the update process of the edge model is performed. Since the parameters of the above process are solved according to a variant of gradient descent, a proof is required:

$$E\|v_k(t_1) - \omega^*\|^2 \leq I(v_k(t_k) - \omega^*) \quad (29)$$

As in Eq. 29, the distance between the parameters of the current iteration and the optimal parameters ω^* is less than the distance between the parameters of the previous iteration and the optimal parameters ω^* , which means that the upper bound of the distance between the parameters of the current iteration and the optimal parameters is decreasing.

There are some assumptions on the functions F_1, \dots, F_K for the algorithm as the premise of convergence analysis:

- 1) As in Eq. 30, F_1, \dots, F_K are all L -smooth: for all v and ω ,

$$F_k(v) \leq F_k(\omega) + (v - \omega)^T \nabla F_k(\omega) + \frac{L}{2} \|v - \omega\|^2 \quad (30)$$

- 2) As in Eq. 31, F_1, \dots, F_K are all μ -convex: for all v and ω ,

$$F_k(v) \geq F_k(\omega) + (v - \omega)^T \nabla F_k(\omega) + \frac{\mu}{2} \|v - \omega\|^2 \quad (31)$$

- 3) The variance of the stochastic gradient is bounded as in Eq. 32:

$$E\|\nabla F_k(v) - \nabla F_k(\omega)\| \leq \sigma_k^2 \quad (32)$$

- 4) The expectation of the 2-Norm of the stochastic gradient is consistently bounded as in Eq. 33:

TABLE 4 Comparison of RMSE under different federated learning algorithms.

Federated learning algorithms	epochs = 5		epochs = 10		epochs = 20	
	rounds = 20		rounds = 10		rounds = 5	
	RMSE	Average RMSE	RMSE	Average RMSE	RMSE	Average RMSE
FedAvg	2.685	2.547	2.377	2.362	2.419	2.368
	2.405		2.495		2.228	
	2.479		2.484		2.288	
	2.506		2.269		2.259	
	2.660		2.185		2.646	
FedProx	2.550	2.543	2.267	2.358	2.218	2.365
	2.516		2.336		2.323	
	2.489		2.447		2.388	
	2.505		2.284		2.391	
	2.655		2.456		2.505	
FedPer	2.690	2.546	2.302	2.357	2.227	2.372
	2.495		2.406		2.203	
	2.411		2.333		2.237	
	2.553		2.309		2.399	
	2.581		2.435		2.794	
DAAFed	2.447	2.460	2.282	2.313	2.398	2.334
	2.317		2.294		2.255	
	2.386		2.316		2.429	
	2.414		2.308		2.267	
	2.736		2.365		2.321	

$$E\|\nabla F_k(\mathbf{v})\|_2^2 \leq G^2 \quad (33)$$

Meanwhile, we measure the system heterogeneity by Γ as in Eq. 34, which denotes the difference between the optimal value of the global objective function and the weighted local objective function. Γ converges to 0 when the data are distributed iid independently and identically.

$$\Gamma = F^* - \sum_k \lambda_k F_k^* \quad (34)$$

Let $\mathbf{v}' = \mathbf{v}_k(t_1)$ and $\mathbf{v} = \mathbf{v}_k(t_k)$, then

$$\begin{aligned} & E\|\mathbf{v}' - \omega^*\|^2 \\ &= E\|\mathbf{v} - \omega^* - \eta_1 \nabla F_k(\mathbf{v}) \times [\lambda_k \nabla F_k(\mathbf{v}) + \mathbf{u}_k(\mathbf{v} - \omega)]\|^2 \\ &= E(\|\mathbf{v} - \omega^*\|^2 + \eta_1^2 [\lambda_k \nabla F_k(\mathbf{v}) + \mathbf{u}_k(\mathbf{v} - \omega)]^2 - 2\eta_1 \times \\ &\quad < \mathbf{v} - \omega^*, [\lambda_k \nabla F_k(\mathbf{v}) + \mathbf{u}_k(\mathbf{v} - \omega)] \nabla F_k(\mathbf{v}) >) \end{aligned} \quad (35)$$

in which

$$\begin{aligned} & -2\eta_1 < \mathbf{v} - \omega^*, [\lambda_k \nabla F_k(\mathbf{v}) + \mathbf{u}_k(\mathbf{v} - \omega)] \nabla F_k(\mathbf{v}) > \\ &= -2\eta_1 < \mathbf{v} - \omega + \omega - \omega^*, [\lambda_k \nabla F_k(\mathbf{v}) + \mathbf{u}_k(\mathbf{v} - \omega)] \nabla F_k(\mathbf{v}) > \\ &\leq -2\eta_1 \left[\mathbf{u}_k T_k + \frac{\mu}{2} \|\mathbf{v} - \omega^*\|^2 \right] + 6\Gamma L \eta_1^2 - \lambda_k T_k \end{aligned} \quad (36)$$

$$\begin{aligned} & [\lambda_k \nabla F_k(\mathbf{v}) + \mathbf{u}_k(\mathbf{v} - \omega)]^2 \\ &= \lambda_k^2 \nabla F_k(\mathbf{v})^2 + \mathbf{u}_k^2 \|\mathbf{v} - \omega\|^2 + 2\lambda_k \mu_k \nabla F_k(\mathbf{v})(\mathbf{v} - \omega) \\ &\leq \lambda_k^2 G^2 + \mathbf{u}_k^2 T_k + 2\lambda_k \mu_k \frac{\mu}{2} T_k \end{aligned} \quad (37)$$

In summary, the original equation can be converted as follow:

$$E\|\mathbf{v}' - \omega^*\|^2 \leq (1 - \eta_1 \mu) E\|\mathbf{v} - \omega^*\|^2 + \eta_1^2 B - (\lambda_k + 2\eta_1 \mu_k) T_k \quad (38)$$

$$B = 6\Gamma L + \lambda_k^2 G^2 + \mathbf{u}_k^2 T_k + \lambda_k \mu_k \mu T_k \quad (39)$$

As in Eqs 35–39, this convergence analysis is formally similar to that of FedAvg. However the equation in this paper subtracts one more positive term, which show that the iterative process is convergent.

4.3 Decentralization mechanism

Considering the demand for distributed power time-series data prediction and the disadvantages of centralized processing, the approach of applying blockchain technology in this case is proposed in this paper in order to decentralize prediction,

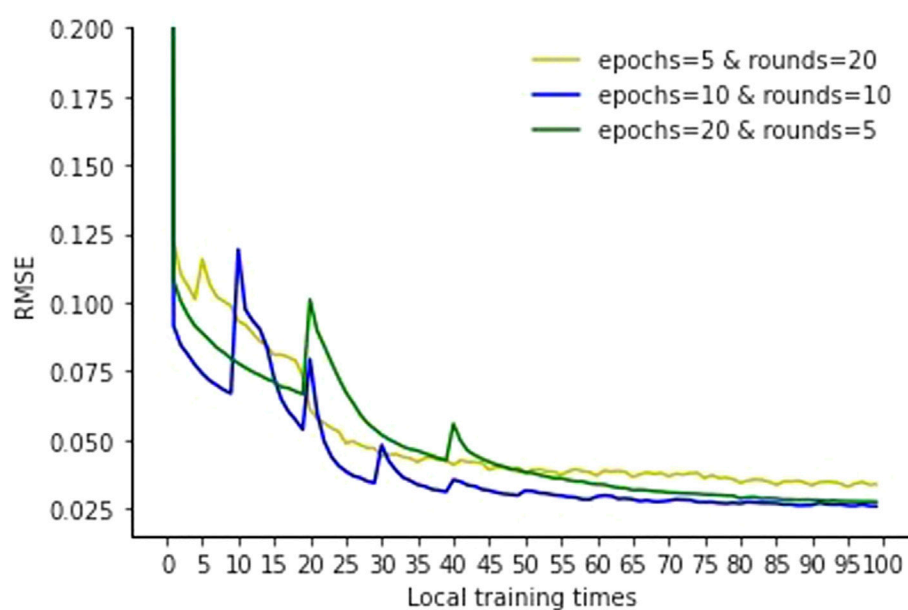


FIGURE 7
Comparison of iteration curves under different iteration parameter settings.

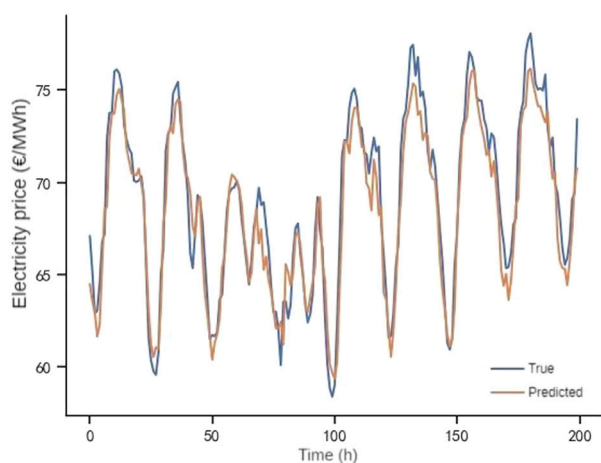


FIGURE 8
Predicted results of FedAvg.

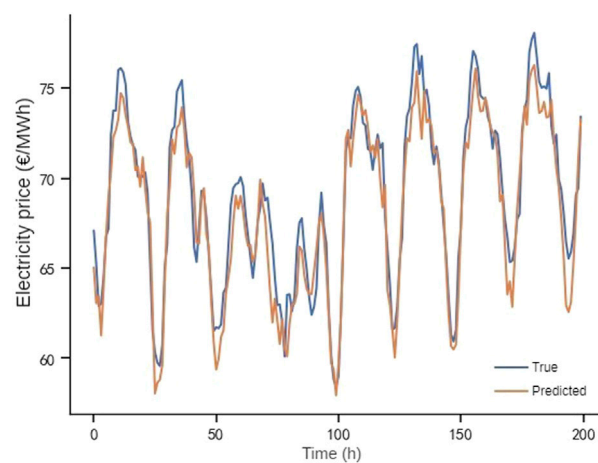


FIGURE 9
Predicted results of FedProx.

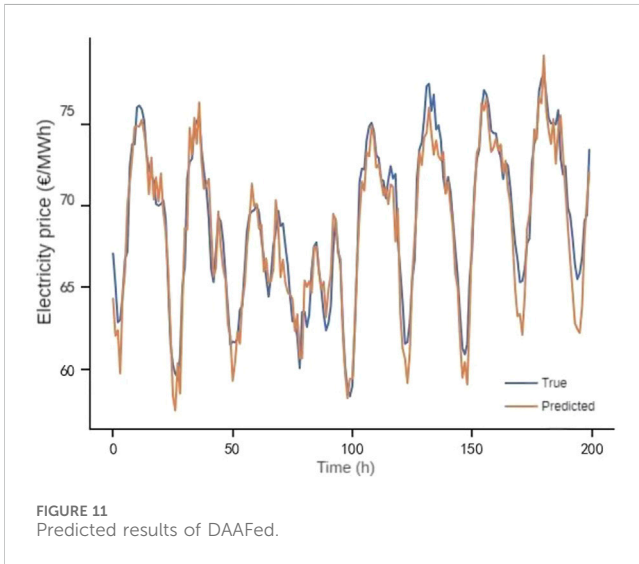
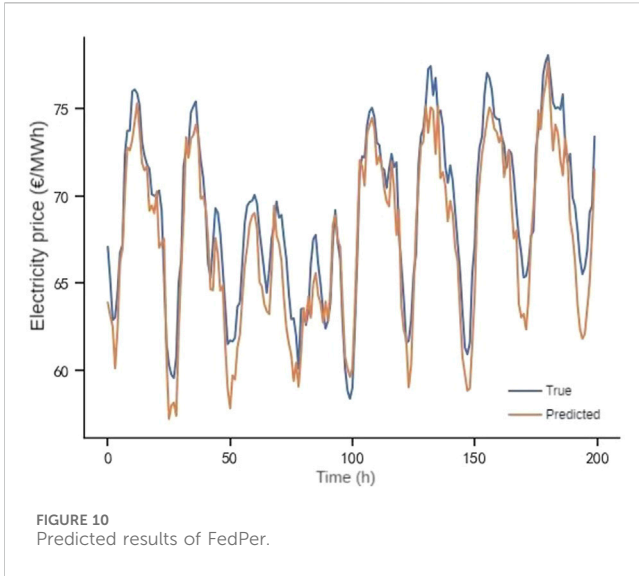
leveraging its global data ledger sharing and non-tamperability, which can further ensure the trust, security, transparency and traceability of data in the power industry.

In this paper, the federated learning global model parameter updating process described in the previous subsection is implemented by placing it on a blockchain. The blockchain guarantees the operation of a shared and trusted distributed ledger on a peer-to-peer network, which in the prediction scenario of this paper is the shared set of global model parameters blocks. The global model used for prediction is placed on the blockchain for updating so that it is not under the control of any single node, but each regional node has equal rights to

validate and access the global model in the blockchain in such a way that it not only avoids a single point of failure, but also protects against data attacks in terms of security.

The workflow of the communication layer of the decentralized federated learning framework used in this paper broadly includes the processes of block generation, consensus verification and global maintenance. The decentralized implementation of the prediction framework includes the following:

- 1) Setting permissions for each regional participant, only participants with permissions for electricity data prediction can jointly train the federation model, and



new nodes can join and receive the current federation model only if they meet the permission conditions when accessing.

- 2) After each edge node completes a round of local training, it packages and broadcasts the privacy protected local model parameters and other data to the network, which includes the local task start time t_k , the local task completion time t_1 , the local model parameters $v_k(t_1)$, and the task completion situation $\varphi_k(t_1)$ and other data to be recorded.

The local update is completed to form a data packet:

$$TX_0 = \{t_k, t_1, v_k(t_1), \dots\} \quad (40)$$

- 3) Each edge node collects the latest packet TX_0 broadcast in the network, completes the update of the global model by block mining, calculates the global model aggregation $\omega(t_1)$, and packages the data before and after the aggregation.

The global aggregation is completed to form a data packet:

$$TX_1 = \{\varphi_k(t_1), \omega(t_1), \dots\} \quad (41)$$

- 4) The node that completes block mining packages the above packets to form a block, and broadcasts it to the network of nodes waiting for verification. The block consists of a block body and a block header, where the block body contains the data before and after aggregation, and the block header contains the predecessor hash and the root hash, and the specific process of block generation includes:

Local updating and global aggregation of data packets together form the block body:

$$Body_N = \{TX\} = \{TX_0, TX_1, \dots\} \quad (42)$$

Hash algorithm achieves the irreversible mapping from plaintext to ciphertext, this paper uses SHA256 to calculate the hash value. The successor hash is the hash value of the previous block and the root hash is generated based on the hash value of the packet $\{TX\}$:

$$PrevHash = Hash(Header_{N-1}) \quad (43)$$

$$RootHash = MerkleTree(\{TX\}) \quad (44)$$

The successor hash and the root hash together form the block header of the block:

$$Header_N = \{PrevHash, RootHash\} \quad (45)$$

The block header and the block body together form the block:

$$Block_N = \{Header_N, Body_N\} \quad (46)$$

- 5) After receiving blocks, all nodes in the network collaboratively verify updates to the global model. The fastest node to complete verification within each region sends its signature to the other nodes for verification. Once all nodes have reached a consensus, the block is added to the global model block set. The most recent global model, along with signatures and timestamps, form new blocks and attach to the blockchain. Finally, each node receives a copy of the most recent global model block to synchronise the model. Additionally, all nodes collectively store and maintain the block set of the global model.

At the same time, as shown in Eqs 47, 48, the node that successfully mines a new block can obtain the block reward $BlockReward(k)$, which in turn encourages the nodes to jointly maintain the update of the global model. Meanwhile, the node under the proof-of-work mechanism can obtain the reward $f(S_k)$ proportional to the size of the sample data, which in turn encourages the participants to contribute the data and the model. After joining the prediction federation, participants can obtain additional financial benefits by earning rewards while obtaining accurate prediction models.

$$Reward(k) = BlockReward(k) + f(S_k) \quad (47)$$

$$BlockReward(k) = \begin{cases} BaseReward, & \text{successful mining} \\ 0, & \text{others} \end{cases} \quad (48)$$

4.4 The overall process

In summary, the pseudo-code of the decentralized asynchronous adaptive federated learning algorithmic process is shown in [Table 1](#). Combining the advantages of federated learning and blockchain, the framework achieves privacy protection for local multi-source data while removing the dependence on a centralized central server. The approach improves training reliability and reduces communication costs, while ensuring the scalability of the edge node network and reducing the risk of single point of failure.

5 Experiments and results

5.1 Data and setup

The proposed method is implemented on a high performance server equipped with Intel Xeon Gold 6136 and NVIDIA TITAN XP. The Python libraries, including TensorFlow and Keras are used to build our model. The experimental data was selected as a dataset of historical data from 1st January 2015 to 31st December 2018 for five regions of the Spanish electricity market, which consists of time-series data on electricity price, generation, load, weather, and date information, sampled every 1 h, where the price of electricity is used as a label for the prediction. Missing values and outliers in the series are filled and replaced using linear interpolation, normalization is used before prediction training to prevent the effect of different magnitudes, and inverse normalization is used for prediction output to restore the relevance of the output data.

5.2 Analysis of the overall implementation effect

First of all, to show the overall implementation effect of the proposed framework in this paper, the data set is divided into training set, validation set and test set for prediction training, the RMSE of the training set and validation set in the iterative process is recorded, and finally the prediction of the test set is performed and the result is compared with the real value, and the partial tariff prediction results and RMSE iteration curves of the nodes are shown in [Figure 3](#).

To verify the improvement of the prediction performance of the proposed framework in this paper, the distributed regional tariff prediction model and the centralized tariff prediction model are compared and analyzed, and the prediction results of the two are shown in [Figure 4](#).

As shown in [Figure 4](#), the predicted and actual values of the two prediction models are relatively close to each other, and the predicted value of the distributed model is slightly different from the actual value, because the samples used for training in the centralized model are the data of all nodes, while the samples used for training in each node in the distributed model are only the local data, but the distributed model strengthens the correlation between nodes by protecting the transmission of the model parameters and overcomes the phenomenon of data islanding.

The prediction performances of the distributed and centralized modes are shown in [Table 2](#), except for the prediction accuracy, the

distributed mode mostly outperforms the centralized mode. In terms of running time, since the computational tasks are assigned to each edge node in the distributed mode, and only local data is used for single-point training, the computational volume of a single computational unit is greatly reduced, and the single-point training time in the distributed mode is less than that in the centralized mode. In terms of data transmission, only the model parameters are transmitted in the distributed mode, which protects the data of each node and does not reveal the private data such as power generation information and user habits, and “53601” represents the sum of the elements of the weight matrix and bias vector of the prediction model. The centralized mode requires the transmission of all sample data, which has the risk of data leakage, where “ $5 \times 35064 \times 33$ ” means that there are 35,064 sample data in each of the five nodes, and each sample contains 33 dimensions of data. Therefore, the distributed model can effectively reduce the amount of data transmitted in a single transmission and provide a certain level of security to protect the privacy of electricity.

5.3 Comparative analysis of edge node prediction models

Choosing a more appropriate model for distributed prediction at the edge nodes can effectively improve the overall prediction accuracy. To improve the implementation effect of the method proposed in this paper, LSTM model, CNN model and CNN-LSTM model are used at the edge nodes to predict the electricity price for the same data set. The model parameters are adjusted based on the method proposed in the related literature, and the prediction performance under the three models is compared and analyzed, as shown in [Figure 5](#). The RMSE is calculated again after converting the model output into real size, and the results of the RMSE of each model evaluation index are shown in [Table 3](#).

From [Figures 5, 6](#) and [Table 3](#), it can be seen that the overall prediction value of CNN-LSTM is closest to the real value, and the efficiency of the training iteration is also higher. The main reason is that CNN-LSTM combines the advantages of CNN and LSTM, capturing both local and global temporal features at different time scales through sliding window and loop structure, while the feature of parameter sharing of convolutional layer reduces the number of parameters of LSTM network and improves the computational efficiency. Therefore, choosing CNN-LSTM model for prediction at edge nodes can have good accuracy and efficiency.

5.4 Comparative analysis of federated learning algorithms

To verify that the DAAFed algorithm proposed in this paper has better prediction performance, FedAvg, FedProx ([Li et al., 2020b](#)), FedPer ([Arivazhagan et al., 2019](#)) and DAAFed are used to predict the electricity price on the same dataset, and the prediction performance of the above algorithms are compared and analysed. FedProx adds regularization terms to constrain the similarity between the global model and the local model to prevent overfitting ([Li et al., 2020a](#)). Each node model in FedPer has its

own personalization layer that preserves the characteristics of the node data (Arivazhagan et al., 2019).

The RMSE and average RMSE for each node using the above federated learning algorithms with different iteration settings are shown in Table 4. FedProx adds a regularization term to limit the similarity between the global model and the local model and prevent overfitting; and each node has its own personalization layer in each node's model in FedPer, which is used to preserve the node data's own characteristics. Table 4 shows that the prediction performance of these two improved algorithms is similar to that of FedAvg. The DAAFed algorithm used in this paper has a smaller error compared to the other three algorithms, because the asynchronous mechanism used in the aggregation of the global model allows the nodes to update the model based on the latest global model, and at the same time, the step size is adaptively adjusted in the updating of the node's local model, so the model learns to the better parameter faster, and thus the overall performance is better. Therefore, the overall performance is better.

From Table 4, it can be seen that setting different local iteration epochs and global aggregation rounds has different prediction effects, and choosing the appropriate settings can fully exploit the advantages of the federated learning framework.

Figure 7 shows the iteration curves under different settings. It shows that when epochs = 10 and rounds = 10 there is better iteration efficiency than the other two settings. When epochs = 5 there is a slower decrease in the iteration curve because when the number of local iterations is small, the local model of each node does not have enough performance and cannot make full use of the local data. When epochs = 20 there is a slow decrease in the iteration curve before global aggregation, which means that a higher number of local iterations leads to a slower decrease of the iteration curve before global aggregation. This means that if the number of local iterations is high, it can lead to overfitting, which affects the ability of global model generalization, so the setting of epochs = 10 and rounds = 10 can give better federated learning results.

Based on this setting, the prediction results of a node under each of the above algorithms are shown in Figures 8–11, from which it can be seen that the predicted value of DAAFed is closer to the real value compared to several other federated learning algorithms.

6 Conclusion

In order to solve the problems of high computational cost and low data security under the traditional centralized electricity data forecasting method, this paper proposes a distributed regional electricity market electricity price forecasting method based on decentralized adaptive federated learning. The following conclusions are drawn from the analysis of the arithmetic example of the Spanish electricity market.

Compared to the conventional centralized prediction method, the distributed data prediction method reduces the amount of prediction calculation and data transmission, and approaches the centralized prediction accuracy in the implementation process, while protecting the local data in each region. Precise selection of prediction models on each node and appropriate configuration of the number of local iterations and global update rounds can be highly effective in enhancing the accuracy and efficiency of

prediction. Compared to the Federated Learning algorithm, DAAFed enhances prediction accuracy by introducing an asynchronous adaptive mechanism. The validity of this proposed method is verified.

At present, privacy protection is becoming increasingly important in the power industry, which brings new challenges to distributed electricity price prediction. Therefore, how to further consider the protection effect of privacy protection mechanism and communication efficiency optimization will be the focus of the next research work.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

QL: Writing—original draft, Writing—review and editing, Conceptualization, Methodology, Validation. DL: Software, Supervision, Writing—review and editing. HC: Funding acquisition, Resources, Writing—review and editing. XiL: Formal Analysis, Project administration, Writing—review and editing. XuL: Data curation, Methodology, Visualization, Writing—review and editing. WC: Conceptualization, Investigation, Writing—review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This research is supported by the State Grid Information and Telecommunication Group scientific and technological innovation projects “Research on Power Digital Space Technology System and Key Technologies” (Grant No. SGIT0000XMJS2310456).

Conflict of interest

Authors QL, DL, XiL, and WC were employed by State Grid Information & Telecommunication Group Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Ahmed, F. K., Hussain, J. S., Attique, M. K., Sharif, M., Tariq, U., and Kadry, S. (2022). A review on federated learning towards image processing. *Comput. Electr. Eng.* 99, 107818. doi:10.1016/j.compeleceng.2022.107818
- Ali, H., Chandren, R., Kamrul, M., Paw, J. K. S., and Singh, M. J. (2023). Big data analytics using cloud computing based frameworks for power management systems: status, constraints, and future recommendations. *Sensors* 23 (6), 2952. doi:10.3390/s23062952
- Antal, M., Mihailescu, V., Cioara, T., and Anghel, I. (2022). Blockchain-based distributed federated learning in smart grid. *Mathematics* 10 (23), 4499. doi:10.3390/math10234499
- Arivazhagan, M., Aggarwal, V., and Singh, A. (2019). Federated learning with personalization layers. ArXiv:1912.00818. Available at: <https://doi.org/10.48550/arXiv.1912.00818>.
- Badhiye, S., Chatur, P., and Raghuwanshi, M. (2022). Time series forecasting using range regression automata. *International journal of uncertainty. Fuzziness Knowledge-Based Syst.* 30 (06), 1035–1063. doi:10.1142/S0218488522500325
- Chen, Y., Lu, W., Qin, X., Wang, J., and Xie, X. (2023). MetaFed: federated learning among federations with cyclic knowledge distillation for personalized healthcare. *IEEE Trans. neural Netw. Learn. Syst.* 2023, 1–12. doi:10.1109/tnnls.2023.3297103
- Frizzo, S., Oriol, L., Cocco, V., and Coelho, L. d. S. (2023). Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices. *Energies* 16 (3), 1371. doi:10.3390/en16031371
- He, W., and Zhao, L. (2022). Application of federated learning algorithm based on K-means in electric power data. *J. New Media* 4 (4), 191–203. doi:10.32604/jnm.2022.032994
- Husnoo, M., Anwar, A., Reda, H., Hosseinzadeh, N., Islam, S., Mahmood, A., et al. (2023). FedDiSC: a computation-efficient federated learning framework for power systems disturbance and cyber attack discrimination. *Energy AI* 14, 100271. doi:10.1016/j.egyai.2023.100271
- Li, T., Anit, K., Ameet, T., and Smith, V. (2020a). Federated learning: challenges, methods, and future directions. *IEEE Signal Process. Mag.* 37 (3), 50–60. doi:10.1109/msp.2020.2975749
- Li, T., Sahu, A., and Zaheer, M. (2020b). Federated optimization in heterogeneous networks. *Proc. Mach. Learn. Syst.* 2020 (2), 429–450. doi:10.48550/arXiv.1812.06127
- Li, Y., Wang, R., Li, Y., Zhang, M., and Long, C. (2023a). Wind power forecasting considering data privacy protection: a federated deep reinforcement learning approach. *Appl. Energy* 329, 120291. doi:10.1016/j.apenergy.2022.120291
- Li, Y., Yan, Y., Liu, Z., Yin, C., Zhang, J., and Zhang, Z. (2023b). A federated learning method based on blockchain and cluster training. *Electronics* 12 (19), 4014. doi:10.3390/electronics12194014
- Liu, Y., Wang, G., Guo, W., Zhang, Y., Dong, W., Wei, G., et al. (2021). Power data mining in smart grid environment. *J. Intelligent Fuzzy Syst.* 40 (2), 3169–3175. doi:10.3233/jifs-189355
- McMahan, H., Moore, E., and Ramage, D. (2017). “Communication-efficient learning of deep networks from decentralized data,” in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 1273–1282.
- Mohamed, A., Ibrahim, A., Hossam, H., and Karam, S. (2023). Towards efficient and trustworthy pandemic diagnosis in smart cities: a blockchain-based federated learning approach. *Mathematics* 11 (14), 3093. doi:10.3390/math11143093
- Qi, T., Chen, L., Li, G., Li, Y., and Wang, C. (2023). FedAGCN: a traffic flow prediction framework based on federated learning and Asynchronous Graph Convolutional Network. *Appl. Soft Comput.* 138, 110175. doi:10.1016/j.asoc.2023.110175
- Qu, Z., Li, Y., Liu, B., Gupta, D., and Tiwari, P. (2023). DTQFL: a digital twin-assisted quantum federated learning algorithm for intelligent diagnosis in 5G mobile network. *IEEE J. Biomed. Health Inf.* 2023, 1–10. doi:10.1109/jbhi.2023.3303401
- Wang, Z., Zhou, M., Zhao, Y., Zhang, F., Wang, J., Qian, B., et al. (2023). Electrical power edge-end interaction modeling with time series label noise learning. *Electronics* 12 (18), 3987. doi:10.3390/electronics12183987
- Wu, X., Liang, Z., and Wang, J. (2020). FedMed: a federated learning framework for language modeling. *Sensors* 14, 4048. doi:10.3390/s20144048
- Yu, H., Yang, S., and Zhu, S. (2019). “Parallel restarted SGD with faster convergence and less communication: demystifying why model averaging works for deep learning,” in Proceedings of the AAAI Conference on Artificial Intelligence, 5693–5700.
- Zeng, M., Wang, X., Pan, W., and Zhou, P. (2023). Heterogeneous training intensity for federated learning: a deep reinforcement learning approach. *IEEE Trans. Netw. Sci. Eng.* 10 (2), 990–1002. doi:10.1109/tNSE.2022.3225444
- Zhang, X., Zhong, C., Zhang, J., and Wang, T. (2023). Robust recurrent neural networks for time series forecasting. *Neurocomputing* 526, 143–157. doi:10.1016/j.neucom.2023.01.037



OPEN ACCESS

EDITED BY

Bo Wang,
Wuhan University, China

REVIEWED BY

Jun-Wei Zha,
University of Science and Technology Beijing,
China
Qu Zhou,
Southwest University, China

*CORRESPONDENCE

Zhengping Liang,
✉ zpliang188@163.com
Yiyi Zhang,
✉ yiyizhang@gxu.edu.cn

RECEIVED 02 December 2023

ACCEPTED 22 December 2023

PUBLISHED 12 January 2024

CITATION

Liang Z, Sun Y, Cheng H, Zhang N, Li B, Liu Y, Fang Y and Zhang Y (2024), A comprehensive evaluation method for moisture and aging state of oil-immersed paper bushings based on dielectric modulus and K-nearest neighbor algorithm.
Front. Energy Res. 11:1348433.
doi: 10.3389/fenrg.2023.1348433

COPYRIGHT

© 2024 Liang, Sun, Cheng, Zhang, Li, Liu, Fang and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A comprehensive evaluation method for moisture and aging state of oil-immersed paper bushings based on dielectric modulus and K-nearest neighbor algorithm

Zhengping Liang^{1*}, Yongbin Sun¹, Hao Cheng¹, Na Zhang¹, Bo Li², Yang Liu¹, Yan Fang¹ and Yiyi Zhang^{2*}

¹North China Power Engineering Co., Ltd. of China Power Engineering Consulting Group, Beijing, China,
²Guangxi Power Transmission and Distribution Network Lightning Protection Engineering Technology Research Center, Guangxi University, Nanning, China

The insulation performance of oil-immersed paper bushings is prone to deteriorate, primarily due to moisture intrusion and thermal aging. The frequency domain spectroscopy (FDS) method is commonly employed to assess the insulation condition of the bushing. However, identifying and extracting relaxation polarization information from the low-frequency region of the FDS curve can be challenging, and there is little research about the condition evaluation under the combined effects of aging states and moisture content. To address this issue, this article uses the Taylor formula mathematical model to extract characteristic parameters from the dielectric modulus curve of OIP bushings and uses the KNN algorithm to achieve the evaluation of aging and moisture status. Then, the effectiveness and accuracy of the proposed method are validated on three field OIP bushings. The results demonstrate that the evaluation accuracy of the proposed method exceeds 83%, which has significant advantages compared to other classification algorithms. The innovation of this article lies in extracting new feature parameters and combining them with intelligent classification algorithms to evaluate the moisture and aging state of the bushing.

KEYWORDS

aging state, dielectric modulus, frequency domain spectroscopy (FDS), K-nearest neighbor algorithm, moisture content

1 Introduction

Oil-impregnated paper (OIP) bushings are extensively employed in power systems and hold the dominant market share, encompassing approximately 80% of the global bushing market (Bouaicha et al., 2009; Jyothi and Ramu, 2012). Once the bushing failure occurs, it may directly or indirectly cause huge economic losses and affect the stable operation of the power system (Wang et al., 2019; Liao et al., 2021). Statistics indicate that bushing faults account for a significant portion, ranging from 30% to 45%, of various power accidents (Yang et al., 2019; Liu et al., 2020a). Additionally, more than 37% of transformer explosions

and fire incidents can be attributed to bushing faults, with a rising trend (Teng et al., 2022; Su et al., 2023). Among numerous factors influencing bushing performance, aging and moisture have the most substantial impact on the insulation (Teng et al., 2022). Therefore, deterioration in this context typically refers to the effects of aging and moisture intrusion on the oil-paper insulation of OIP bushings (Li et al., 2021; Zhou et al., 2021; Akbari et al., 2023). The degree of polymerization (DP) is usually considered the most important indicator for evaluating the aging degree of insulation papers (Li et al., 2020). Consequently, conducting a comprehensive evaluation of the bushing's moisture condition and dielectric polarization (DP) holds immense significance.

In recent years, frequency domain spectroscopy (FDS) technology has been widely applied in the assessment of bushing conditions (Saha and Purkait, 2004; Fofana and Hadjadj, 2016; Zhang et al., 2019a; Xie et al., 2019). Compared with other dielectric response methods, it has advantages such as practicality, non-destructive, and rich insulation information (Buchacz et al., 2017; Netaworldjournal, 2018).

For the status evaluation of oil-paper insulation (Linhjell et al., 2007) pointed out that an increase in the moisture content of the insulating pressboard can cause the complex relative dielectric constant curve to shift towards the coordinate axis at high frequencies (Jadav et al., 2014) pointed out that the complex relative permittivity curves are more sensitive to the moisture content of the pressboard, while the sensitivity of the curves to aging is relatively small (Zaengl, 2003). Establish a quantitative relationship between the moisture content of insulation paper and the $\tan \delta$ curve (Liu et al., 2021) extracted FDS test feature values from insulating paper with different levels of moisture content, and applied the GA-SVM (Support Vector Machine Based on Genetic Algorithm Optimization) algorithm to diagnose moisture content under laboratory and on-site conditions (Poovamma et al., 2008). found that the dielectric loss and the imaginary part of the complex relative permittivity of the insulating paperboard increases with the increase of aging degree, and the peak of the $\tan \delta$ curve appears in the low-frequency region (Zhang et al., 2019b) obtained a new characteristic quantity based on the Davidson Cole model and the complex relative permittivity curve, and proposed a new quantitative evaluation method for the aging state of oil-paper insulation (Ren et al., 2019) reported a quantitative evaluation method for the insulation aging status of transformers based on frequency domain dielectric response test results by taking into account the Arrhenius equation and Ekendam equation.

However, the FDS parameters adopted in the above method is not applicable to the study of the relaxation process of the oil-paper insulation of the transformer bushings in the low-frequency region, which in turn makes the information on the insulation state carried by it ambiguous and leads to errors in the results. In addition, the state assessment of oil-paper insulation by the above methods often only considers the influence of a single factor, in fact, the moisture and aging process is always accompanied in the bushing at the same time, so the assessment of a single state of the bushing may lead to unsatisfactory assessment results.

Compared with the traditional FDS curves, the dielectric modulus has an outstanding advantage in characterizing the dielectric properties of insulating materials, as it carries more comprehensive information about the relaxation of insulating

materials and reduces the influence of electrode polarization and conductivity behavior on the measurement results. In addition, the K-Nearest Neighbor (K-NN) classification algorithm has the advantages of high accuracy, insensitivity to anomalies, and conceptual clarity, and has been widely used in text and image recognition in recent years (Zhang et al., 2017; Islam et al., 2018; Liao et al., 2020). Besides, the K-NN classification algorithm has high flexibility, universality and evaluation ability for the comprehensive evaluation of two dimensions of moisture and aging state of oil-paper insulation (Choi et al., 2009; Rouhafzay and Cretu, 2020).

Therefore, this paper establishes a mathematical model of dielectric modulus and analyzes the influence of aging and moisture on the dielectric modulus curve, and proposes a Taylor's formula mathematical model on the basis of the obtained dielectric modulus curve, which extracts the characteristic parameters of the dielectric modulus curves of different moisture and aging states. Then, combined with the K-NN classification algorithm model, three field bushings with unknown insulation states are evaluated for moisture and aging states (Liu et al., 2020b; Fan et al., 2021). Subsequently, an analysis and comparison with other classification algorithms (Support Vector Machine, Naive Bayes, and Decision Tree) is carried out to analyze the evaluation results and errors of several algorithms in detail, and then to prove the effectiveness and accuracy of the proposed evaluation method.

2 Methodology

2.1 Mathematical model of dielectric modulus

According to (Li et al., 2020), the reciprocal of the complex dielectric constant defines the complex dielectric modulus $M^*(\omega)$. This relationship is expressed in Eqs 1, 2.

$$M^*(\omega) = \frac{1}{\epsilon^*}, \quad (1)$$

$$\epsilon^*(\omega) = \epsilon' - i\epsilon''. \quad (2)$$

Equation 3 depicts the formulations for the real and imaginary parts of the dielectric modulus.

$$\begin{aligned} M'(\omega) &= \frac{\epsilon'}{\epsilon'^2 + \epsilon''^2}, \\ M''(\omega) &= \frac{\epsilon''}{\epsilon'^2 + \epsilon''^2}. \end{aligned} \quad (3)$$

If the Debye relaxation model is introduced, the complex permittivity $\epsilon^*(\omega)$ can then be expressed in Eqs 4, 5 as

$$\epsilon^*(\omega) = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + (\omega\tau)^2}, \quad (4)$$

$$\begin{cases} \epsilon'(\omega) = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + (\omega\tau)^2}, \\ \epsilon''(\omega) = \frac{(\epsilon_s - \epsilon_\infty) \cdot \omega\tau}{1 + (\omega\tau)^2}, \end{cases} \quad (5)$$

where ϵ_s is the static dielectric constant, ϵ_∞ is the dielectric constant as the angular velocity approaches infinity, τ is the relaxation time

constant, and the distribution factor β ($0 < \beta < 1$) is related to the shape of the FDS curve plotted in the complex plane (Li et al., 2020).

The detailed relationship between dielectric modulus and complex dielectric constant is as Eq. 6

$$\frac{1}{\varepsilon^*(\omega)} = \frac{1}{\varepsilon_{\infty} + \frac{\varepsilon_s - \varepsilon_{\infty}}{1 + (i\omega\tau)^\beta}} = \frac{1}{\varepsilon_{\infty}} - \frac{1}{1 + \frac{1 + \frac{\varepsilon_s}{\varepsilon_{\infty}}}{1 + (i\omega\tau)^\beta}} = \frac{1}{\varepsilon_{\infty}} - \frac{\frac{1}{\varepsilon_{\infty}} - \frac{1}{\varepsilon_s}}{1 + (i\omega\tau)^\beta \cdot \frac{\varepsilon_{\infty}}{\varepsilon_s}} \\ = \frac{1}{M_{\infty}} - \frac{M_{\infty} - M_s}{1 + (i\omega\tau_M)^\beta} = M^*(\omega), \quad (6)$$

where $M_{\infty} = 1/\varepsilon_{\infty}$, $M_s = 1/\varepsilon_s$ and $\tau_M = \tau(\varepsilon_{\infty}/\varepsilon_s)^{1/\beta}$. Furthermore, since ε_{∞} is always less than ε_s , it is easy to deduce that τ_M is always less than τ . The above relations indicate that the frequency interval characterizing the relaxation polarization in the complex dielectric modulus profile will move to a higher frequency interval than the dielectric constant. If M_{∞} , M_s , and τ_M are substituted into Eq. 5, the expressions for the real and imaginary parts of $M^*(\omega)$ are obtained as shown in Eq. 7.

$$\begin{cases} M'(\omega) = M_{\infty} + \frac{M_s - M_{\infty}}{1 + (\omega\tau_M)^2}, \\ M''(\omega) = \frac{(M_s - M_{\infty}) \cdot \omega\tau_M}{1 + (\omega\tau_M)^2}, \end{cases} \quad (7)$$

When the effect of conductivity is neglected, in mathematical form the complex dielectric modulus $M^*(\omega)$ can be regarded as the reciprocal of the complex relative permittivity $\varepsilon^*(\omega)$, with the frequency f in the denominator position. The dielectric modulus $M^*(\omega)$ serves as a valuable tool for assessing the insulation condition of transformer or bushing oil-paper structures under alternating electric fields. It overcomes the limitation of complex relative dielectric constant $\varepsilon^*(\omega)$ by avoiding the masking of relaxation polarization information in the low-frequency range. Thus, the dielectric modulus $M^*(\omega)$ effectively captures and characterizes the relaxation polarization properties of the oil-paper insulation, contributing to the evaluation of the insulation status of the bushing.

2.2 Taylor formula mathematical model

Taylor's formula is one of the approximation methods often used in order to study the properties of complex functions, and its core idea is to approximate a complex function by replacing it with a polynomial function, within the function's domain of definition. The process of establishing the mathematical model of Taylor's formula is as follows.

If there exists a function $f(x)$, and when $f(x)$ is derivable at x_0 , then $f(x)$ is constant in the neighborhood of x_0 , as presented in Eq. 8:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0), \quad (8)$$

where $o(x - x_0)$ represents an infinitesimal quantity, in order to make the approximate substitution more precise, Taylor's formula of order n ($n \geq 2$) is generally used in analytical studies, as shown in Eq. 9, and n is referred to as the order of Taylor's formula.

$$f(x) = \frac{f(x_0)}{0!} + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + R_n(x), \quad (9)$$

where x_0 is the point with $(n+1)$ order derivative contained in the interval (a, b) of the $f(x)$.

The mathematical modeling process of Taylor's formula based on the $M''(\omega)$ curve is as follows: let $G(f)$ be the functional expression of the dielectric modulus curve $M''(\omega)$, and $G(f)$ is continuous in the frequency domain (2×10^{-4} Hz to 5×10^3 Hz) and conductible, satisfying the n th-order Taylor's formula as follows.

$$G(f) = \frac{G(f_0)}{0!} + \frac{G'(f_0)}{1!}(f - f_0) + \frac{G''(f_0)}{2!}(f - f_0)^2 + \dots + \frac{G^{(n)}(f_0)}{n!}(f - f_0)^n + R_n(f), \quad (10)$$

where f_0 lies in the frequency domain (2×10^{-4} Hz to 5×10^3 Hz) and f_0 can be any value in its interval. $G^{(n)}(f_0)$ denotes the n th-order derivative of the function $G(f)$ at f_0 and $G^{(n)}(f_0)$ is a constant. In order to facilitate the representation of Eq. 10, the constant symbol Λ_n is introduced as shown in Eq. 11, and furthermore, since the value of the residual term $R_n(f)$ is close to 0, Eq. 10 can then be simplified to Eq. 12.

$$\Lambda_0 = \frac{G(f_0)}{0!}, \Lambda_1 = \frac{G'(f_0)}{1!}, \Lambda_2 = \frac{G''(f_0)}{2!}, \Lambda_3 = \frac{G'''(f_0)}{3!}, \dots, \Lambda_n = \frac{G^{(n)}(f_0)}{n!}, \quad (11)$$

$$G(f) = \Lambda_0 + \Lambda_1(f - f_0) + \Lambda_2(f - f_0)^2 + \Lambda_3(f - f_0)^3 + \dots + \Lambda_n(f - f_0)^n. \quad (12)$$

In order to make the equation more concise and clearer, the constant Ω_n is additionally introduced to denote $\Lambda(n)$ and f_0 . Thus, the function $G(f)$ can be written as Eq. 13 (Li et al., 2020).

$$G(f) = \Omega_0 + \Omega_1 f + \Omega_2 f^2 + \Omega_3 f^3 + \dots + \Omega_n f^n \\ = \sum_{i=0}^n \Omega_i \cdot f^i \quad (i = 1, 2, 3, \dots, n). \quad (13)$$

The shape of the dielectric modulus curve $M''(\omega)$ varies under different moisture and aging states. As a result, the correlation coefficient Ω_n obtained by Taylor's formula changes with the values of moisture content $mc\%$ and degree of polymerization (DP). In order to clearly represent the imaginary part of the dielectric modulus curve, Eq. 13 can be written as Eq. 14:

$$M''(f) = \sum_{i=0}^n \Omega_i(mc\%, DP) \cdot f^i \quad (i = 1, 2, 3, \dots, n) \quad (14)$$

The above analysis shows that Ω_i ($i = 0, 1, \dots, n$) in Eq. 14 has a unique corresponding value in the face of $M''(f)$ curves of different moisture and aging states. In other words, the coefficient Ω_i ($i = 0, 1, \dots, n$) in the mathematical model of Taylor's formula has a one-to-one mapping relationship with the corresponding moisture and aging state, so it can be used as a characteristic parameter to characterize the moisture and aging state of the oil-paper insulation, and Taylor's formula can be used as a powerful tool to extract the characteristic parameter.

2.3 Principle of the KNN algorithm

The classification principle of the K-NN algorithm is based on the core idea of the distance from the sample point to the nearest

neighbor point to be identified (Rouhafzay and Cretu, 2020). Considering that the characteristic parameters $\varphi(mc\%, DP)$ encapsulate both moisture and aging information of the sample, the set of m characteristic parameters φ is denoted as the sample point x . This representation can be expressed using Eq. 15.

$$x = \{\varphi_1(mc\%, DP), \varphi_2(mc\%, DP), \dots, \varphi_m(mc\%, DP)\} \quad (15)$$

The Euclidean distance, denoted as $d(x, x_k)$, represents the measure of distance between the test sample point x and its adjacent point x_k , as shown in Eq. 16.

$$d(x, x_k) = \sqrt{\sum_{r=1}^m [\varphi_r(mc\%, DP) - \varphi_r(mc\%_k, DP)]^2}, \quad (16)$$

The similarity $\lambda_k(x, x_k)$ between x and x_k is defined as the reciprocal of the distance $d(x, x_k)$, as illustrated in Eq. 17. In the weighted K-NN algorithm, the weight value is determined by selecting the distances of the K nearest points surrounding the sample points. The weight of the adjacent point w_k is represented by Eq. 18.

$$\lambda_k(x, x_k) = \frac{1}{d(x, x_k)}, \quad (17)$$

$$w_k = \frac{\lambda_k^2}{\sum_{k=1}^K \lambda_k^2}. \quad (18)$$

Given that the sample points of the p -type are denoted as v_p , and the total number of categories for the sample points is c , the discriminant function for identifying the x point is expressed by Eq. 19.

$$g_p(x) = \sum_{k=1}^N y_k w_k = k_p, p = 1, 2, \dots, c. \quad (19)$$

In this equation, N represents the number of sample points for the p -type, k_p denotes the sum of weights for points in the p -class, and y_k is the discriminant parameter used to determine whether x_k is an adjacent point of x , as indicated in Eq. 20.

$$y_k = \begin{cases} 1 & x_k \text{ is the neighboring point to } x, \\ 0 & x_k \text{ is not a neighboring point of } x, \end{cases} \quad (20)$$

$$G_p(x) = \max[g_p(x)] = \max k_p, p = 1, 2, \dots, c. \quad (21)$$

Once Eq. 21 is satisfied, it signifies that the unrecognized point x belongs to the category v_p .

3 Experiments

3.1 Preparation of OIP bushing experimental model

In this paper, the insulating pressboard with a diameter of 140 mm and a thickness of 0.5 mm and Karamay No. 25 mineral insulating oil are used to prepare the test sample of the bushing. The specific performance parameters of the insulating pressboard and insulating oil are shown in Tables 1, 2, respectively. Firstly, dry the insulation paper and insulation oil separately in a vacuum

immersion tank for 48 h (105°C 50 Pa). Then put the dried insulation paper into the dried insulation oil for 48 h (70°C 50 Pa). Finally, seal the dried oil immersion paper insulation for future experimental use. The experimental pre-treatment process is shown in Figure 1. It is worth mentioning that when conducting aging experiments, moisture intrusion experiments, and dielectric response tests on oil-immersed paper bushings, oil-immersed insulation paper is usually used instead of the actual bushing, which has the advantages of material saving, time-saving, easy operation, and strong repeatability.

After the pretreated experiments, the aging tests were conducted on the pretreated oil-immersed insulating pressboard samples. Firstly, group the preprocessed insulating pressboard according to the expected aging days, number it, and place it in an aging box for aging. The temperature of the aging box is set to 150°C. The temperature only affects the aging rate of the insulation paper, without affecting the aging products and principles. Therefore, to quickly obtain samples, using 150°C as the aging temperature is a relatively time-saving and reasonable choice. Prepare insulation paper with aging days of 0, 1, 3, 7, and 15 days, and store it. Randomly select one sample from each group of insulated pressboard samples with different aging degrees, and the DP test results are shown in Table 3. Due to the identical aging conditions of each group of samples, it is believed that the DP value of this sample can characterize the aging degree of all insulating pressboards under the same aging conditions.

On the basis of aging experiments, conduct moisture intrusion experiments on oil-immersed insulating pressboard samples. Firstly, use a Karl Fischer tester for initial moisture detection, and record its moisture content as $a\%$. At this point, it is assumed that the initial moisture content of all insulating pressboards is $a\%$. Group and number the insulating pressboard that needs to undergo moisture testing. Place the divided insulating pressboard with an initial moisture content of $a\%$ into a precision electronic balance, record its initial mass as m , and then expose the pressboard to air for natural moisture absorption. By controlling different moisture absorption times, obtain a pressboard with different moisture content ($mc\%$). If the planned moisture content inside the pressboard needs to reach $b\%$, then the measurement value of the electronic balance will reach $(1-a\%)/(1-b\%)$ and seal the pressboard in an aging jar at a constant temperature of 45°C for 48 h to ensure uniform distribution of moisture inside the pressboard. A total of 20 sets of oil-immersed insulation paper samples were prepared, as illustrated in Figure 2.

3.2 Classification of moisture and aging status of oil-paper insulation

According to the report in (Sumreder and Muhr, 2010), the average DP of newly manufactured dry insulating pressboard is generally between 1,000 and 1,250, and as the aging degree increases, the DP gradually decreases. The existing standard DL/T 984-2005 (DL, 2005) requires that when the DP of insulation paper is less than 250, it is the end of its lifespan, and it can be considered that its insulation performance has completely deteriorated and no longer

TABLE 1 Parameters of insulating pressboard.

Type	Manufacturer	Shape (mm)	Tensile strength (MPa)	Density
T4 Transformer Insulation Paper	Taizhou Weidmann High Voltage Insulation Co., Ltd	Thickness:0.5	MD = 98	0.96 g/cm3
		diameter:140	CMD = 47	

TABLE 2 Parameters of insulating oil.

Type	Manufacturer	tanδ	Pour point (°C)	Flash point (°C)
Karamay No. 25 mineral insulating oil	Chongqing Chuanrun Petrochemical Co., Ltd	4×10^{-4}	≤-45	135

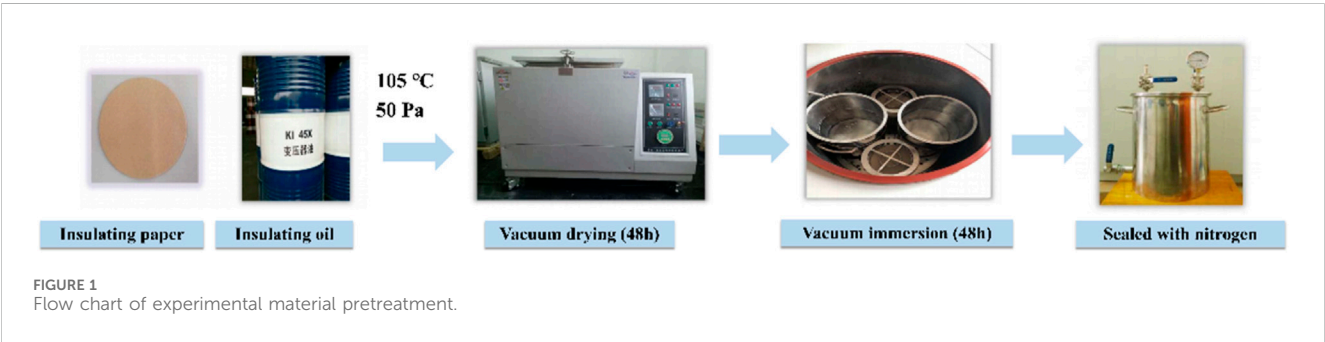


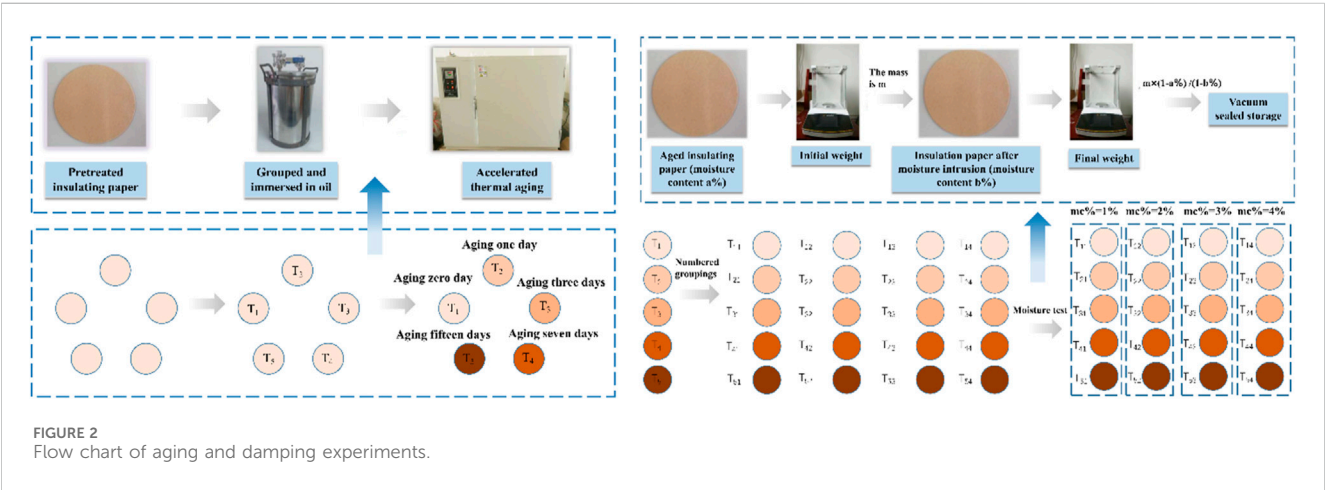
TABLE 3 Polymerization degree values of oil-immersed insulating cardboard samples.

Sample number	T1	T2	T3	T4	T5
Aging day	0	1	3	7	15
DP	1,171	854	674	424	279

meets the insulation requirements. Therefore, within the DP range where the insulating pressboard has insulation properties ($250 \leq DP \leq 1,250$), DP can be divided into five parts, representing five different levels of aging, as shown in Figure 3.

In addition, the technical report published by CIGRE (International Conference on the Great Power Grid) states that

when the moisture content of the insulating pressboard is $0\% < mc\% \leq 1\%$, it is considered dry; When the moisture content of the insulating cardboard is $1\% < mc\% \leq 2\%$, it is considered that the insulating cardboard has been affected by moisture. According to the IEEE standard, a damp insulated pressboard is divided into two stages: moderate moisture and severe moisture. This article believes that when the moisture content of the insulated pressboard is $(2\% < mc\% \leq 3\%)$, it is considered that the insulated pressboard has been moderately damp; When the moisture content of the insulating pressboard is $(mc\% > 3\%)$, it is considered that the insulating pressboard has been heavily damped, as shown in Figure 3. In summary, the oil-immersed insulating pressboard prepared in this article with different levels of moisture and aging will be divided into 20 different states, as shown in Table 4.



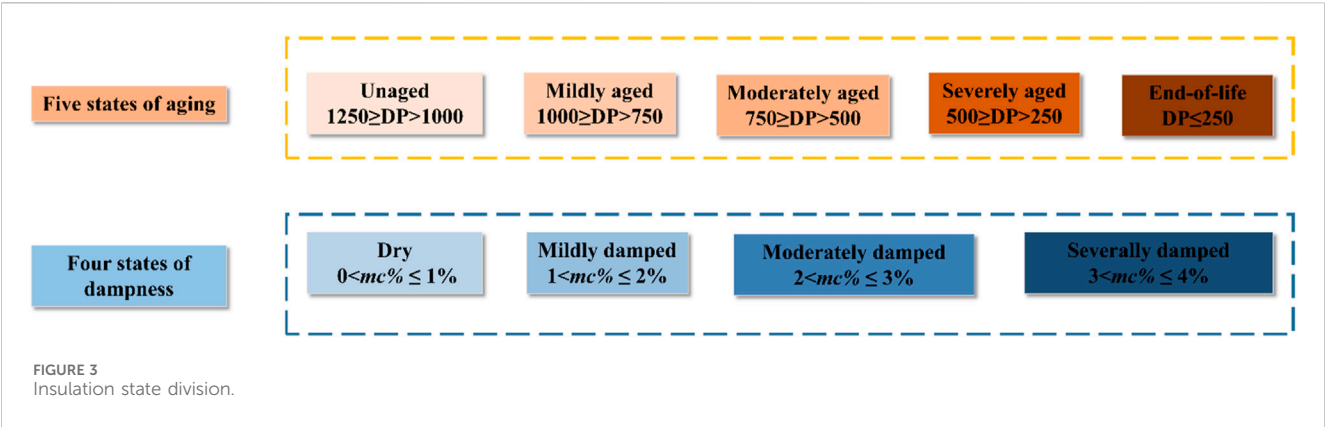


TABLE 4 Classification of moisture and aging conditions of insulated pressboard.

Aging	Moisture content	Status number	Degree of polymerization (DP)	Moisture content (mc%) (%)
Unaging	Dry	A ₁ B ₁	1,000 < DP ≤ 1,250	0% < mc% ≤ 1
	Mildly damp	A ₁ B ₂		1% < mc% ≤ 2
	Moderately damp	A ₁ B ₃		2% < mc% ≤ 3
	Heavily damp	A ₁ B ₄		mc% > 3
Mildly aging	Dry	A ₂ B ₁	750 < DP ≤ 1,000	0% < mc% ≤ 1
	Mildly damp	A ₂ B ₂		1% < mc% ≤ 2
	Moderately damp	A ₂ B ₃		2% < mc% ≤ 3
	Heavily damp	A ₂ B ₄		mc% > 3
Moderate aging	Dry	A ₃ B ₁	500 < DP ≤ 750	0% < mc% ≤ 1
	Mildly damp	A ₃ B ₂		1% < mc% ≤ 2
	Moderately damp	A ₃ B ₃		2% < mc% ≤ 3
	Heavily damp	A ₃ B ₄		mc% > 3
Heavily aging	Dry	A ₄ B ₁	250 < DP ≤ 500	0% < mc% ≤ 1
	Mildly damp	A ₄ B ₂		1% < mc% ≤ 2
	Moderately damp	A ₄ B ₃		2% < mc% ≤ 3
	Heavily damp	A ₄ B ₄		mc% > 3
End of insulation life	Dry	A ₅ B ₁	DP ≤ 250	0% < mc% ≤ 1
	Mildly damp	A ₅ B ₂		1% < mc% ≤ 2
	Moderately damp	A ₅ B ₃		2% < mc% ≤ 3
	Heavily damp	A ₅ B ₄		mc% > 3

4 The influence of moisture and aging states on the dielectric modulus

4.1 Dielectric response test platform construction

In this paper, a DIRANA dielectric response tester and a three-electrode test cell are used to set up a frequency domain dielectric response test platform, the output and test terminals of DIRANA are

connected on both sides of the insulating pressboard, respectively. The test frequency ranged from 2×10^{-4} to 5×10^3 Hz, and the test voltage was set to 200 V AC. Then the FDS tests are performed on insulating paperboards in different aging and moisture states at 45°C, to obtain the real part $\epsilon'(\omega)$ and the imaginary part of the $\epsilon''(\omega)$ of the complex relative permittivity, and then the inverse of the complex permittivity is taken to obtain the real part of the dielectric modulus $M'(\omega)$ and the imaginary part of the permittivity modulus $M''(\omega)$.

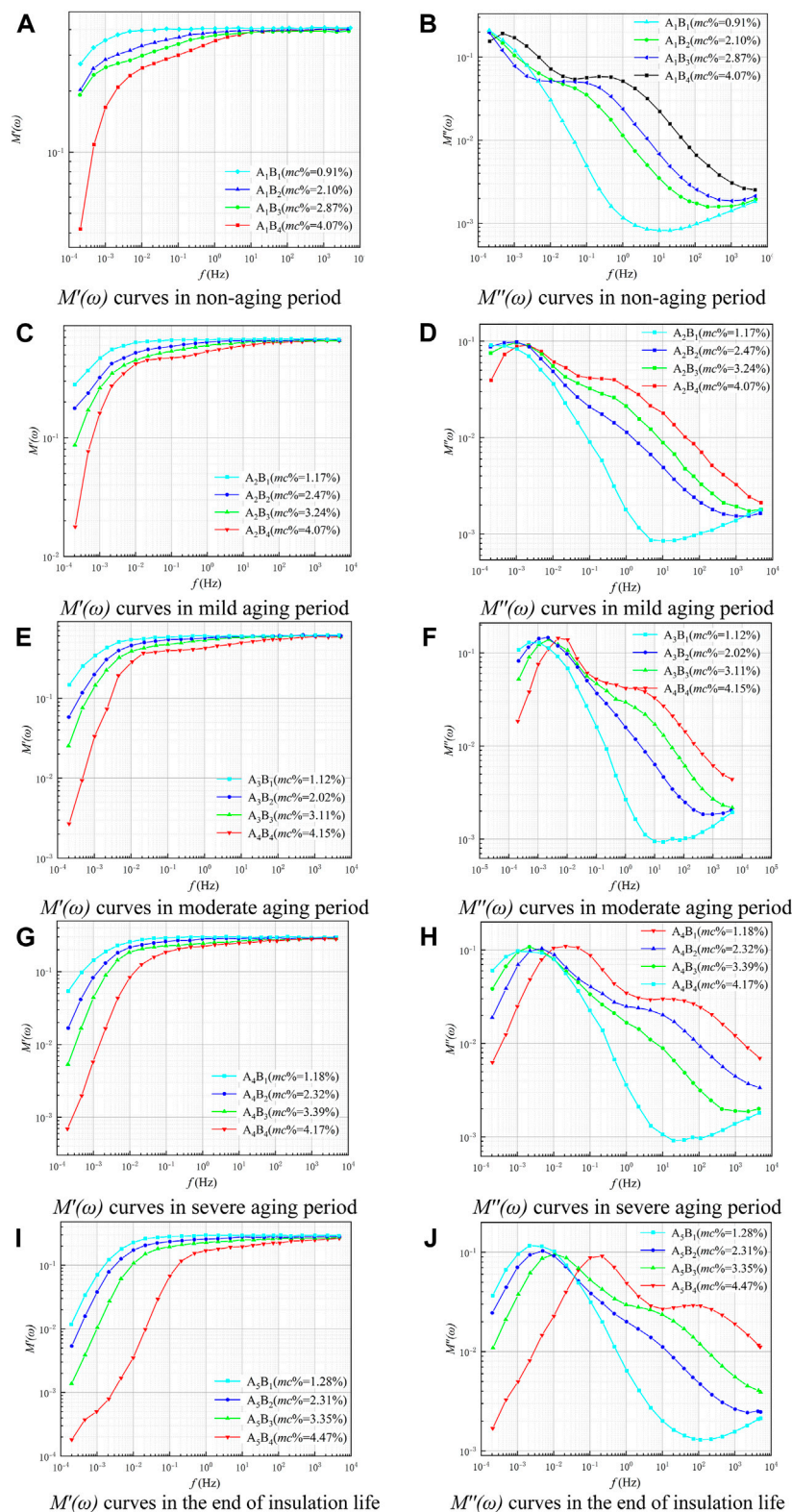


FIGURE 4

$M'(\omega)$ and $M''(\omega)$ curves in different aging periods. (A) $M'(\omega)$ curves in non-aging period. (B) $M''(\omega)$ curves in non-aging period. (C) $M'(\omega)$ curves in mild aging period. (D) $M''(\omega)$ curves in mild aging period. (E) $M'(\omega)$ curves in moderate aging period. (F) $M''(\omega)$ curves in moderate aging period. (G) $M'(\omega)$ curves in severe aging period. (H) $M''(\omega)$ curves in severe aging period. (I) $M'(\omega)$ curves in the end of insulation life. (J) $M''(\omega)$ curves in the end of insulation life.

TABLE 5 Comparison of fitting degree R^2 .

Sample status number	R^2			
	$n = 5$	$n = 6$	$n = 7$	$n = 8$
A ₅ B ₁	0.99975	0.99989	0.99992	0.99992
A ₅ B ₂	0.99715	0.99766	0.99979	0.99987
A ₅ B ₃	0.98541	0.99490	0.99940	0.99955
A ₅ B ₄	0.97222	0.97362	0.99454	0.99586

4.2 Effect of aging and moisture on dielectric modulus

FDS tests were performed on samples with different moisture levels under the same aging degree conditions and the corresponding dielectric modulus graphs were obtained (shown in Figure 4). Under the same aging level, the change of moisture content of oil-impregnated insulating paperboard will have an effect on $M'(\omega)$ in the low and middle-frequency regions, while the high-frequency region (10^2 Hz– 10^4 Hz) is almost unaffected. Comparing oil-impregnated paperboards with the same degree

of aging, an increase in the water content of the paperboard leads to a decrease in the value of $M'(\omega)$ in the low-frequency region of the $M'(\omega)$ curves, while the $M'(\omega)$ curves in the high-frequency region are essentially unchanged under any water content condition. In addition, unlike the imaginary part of the complex relative permittivity, the value of the imaginary part of the dielectric modulus $M''(\omega)$, increases with increasing frequency in the low-frequency region (10^{-4} Hz– 10^{-1} Hz).

4.3 Establishment of the database of dielectric modulus parameters

The Taylor's formula mathematical model based on Eq. 14 is used to extract the relevant characteristic coefficients of the curve. It is worth noting that the error of the extracted parameters depends on the size of the n value in Eq. 14: the larger the value of n , the better the resulting Taylor's formula fitting curve fits the original data, and the smaller the error of the obtained parameters. However, a larger value of n will lead to an increase in computational time. Therefore, this paper selected the imaginary curves of the dielectric modulus of four samples (A₅B₁, A₅B₂, A₅B₃, A₅B₄) at the end-of-life stage of the insulation in the moisture condition to calculate the R^2

TABLE 6 Results of feature parameter extraction.

Status number	$M''(f) = \Omega_0 + \Omega_1 f + \Omega_2 f^2 + \Omega_3 f^3 + \Omega_4 f^4 + \Omega_5 f^5 + \Omega_6 f^6 + \Omega_7 f^7$								R^2
	Ω_0	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	
A ₁ B ₁	−2.9496	−0.3572	0.2081	−0.0026	−0.0157	0.0017	0.0005	−0.0001	0.9996
A ₁ B ₂	−2.0302	−0.5140	−0.0465	0.0599	0.0078	−0.0052	−0.0002	0.0002	0.9999
A ₁ B ₃	−1.7498	−0.4520	−0.1196	0.0566	0.0129	−0.0053	−0.0003	0.0002	0.9999
A ₁ B ₄	−1.4512	−0.1622	−0.1491	−0.0222	0.0204	0.0008	−0.0008	0.00004	0.9991
A ₂ B ₁	−2.7019	−0.6007	0.1873	0.0676	−0.0148	−0.0057	0.0004	0.0002	0.9988
A ₂ B ₂	−1.8907	−0.3126	−0.0323	−0.0228	0.0075	0.0034	−0.0004	−0.0001	0.9995
A ₂ B ₃	−1.6102	−0.2760	−0.0838	−0.0200	0.0117	0.0024	−0.0005	−0.00005	0.9989
A ₂ B ₄	−1.3934	−0.1835	−0.0814	−0.0207	0.0103	0.0012	−0.0006	0.00002	0.9986
A ₃ B ₁	−2.5615	−0.6743	0.1696	0.0742	−0.0139	−0.0057	0.0004	0.0002	0.9992
A ₃ B ₂	−1.7980	−0.3707	−0.0092	−0.0201	0.0028	0.0041	−0.0002	−0.0001	0.9998
A ₃ B ₃	−1.5361	−0.1681	−0.0099	−0.0566	−0.0007	0.0067	−0.0001	−0.0002	0.9997
A ₃ B ₄	−1.3902	−0.0198	0.0412	−0.0948	−0.0103	0.0123	0.0003	−0.0004	0.9950
A ₄ B ₁	−2.3859	−0.8046	0.1463	0.0975	−0.0128	−0.0069	0.0004	0.0002	0.9995
A ₄ B ₂	−1.6704	−0.2736	−0.0006	−0.0507	−0.0009	0.0073	−0.0001	−0.0002	0.9999
A ₄ B ₃	−1.4865	−0.0935	0.0542	−0.0733	−0.0108	0.0093	0.0003	−0.0003	0.9992
A ₄ B ₄	−1.3100	−0.2627	0.146	0.0225	−0.0316	0.0005	0.0012	−0.0001	0.9967
A ₅ B ₁	−2.1434	−0.7143	0.0820	0.0566	−0.0061	−0.0019	0.0001	0.00004	0.9999
A ₅ B ₂	−1.6214	−0.2587	0.0741	−0.0492	−0.0143	0.0078	0.0005	−0.0003	0.9997
A ₅ B ₃	−1.4132	−0.1569	0.0409	−0.0483	−0.0152	0.0080	0.0006	−0.0003	0.9992
A ₅ B ₄	−1.1465	−0.4176	−0.0533	0.1707	−0.0064	−0.0181	0.0003	0.0006	0.9923

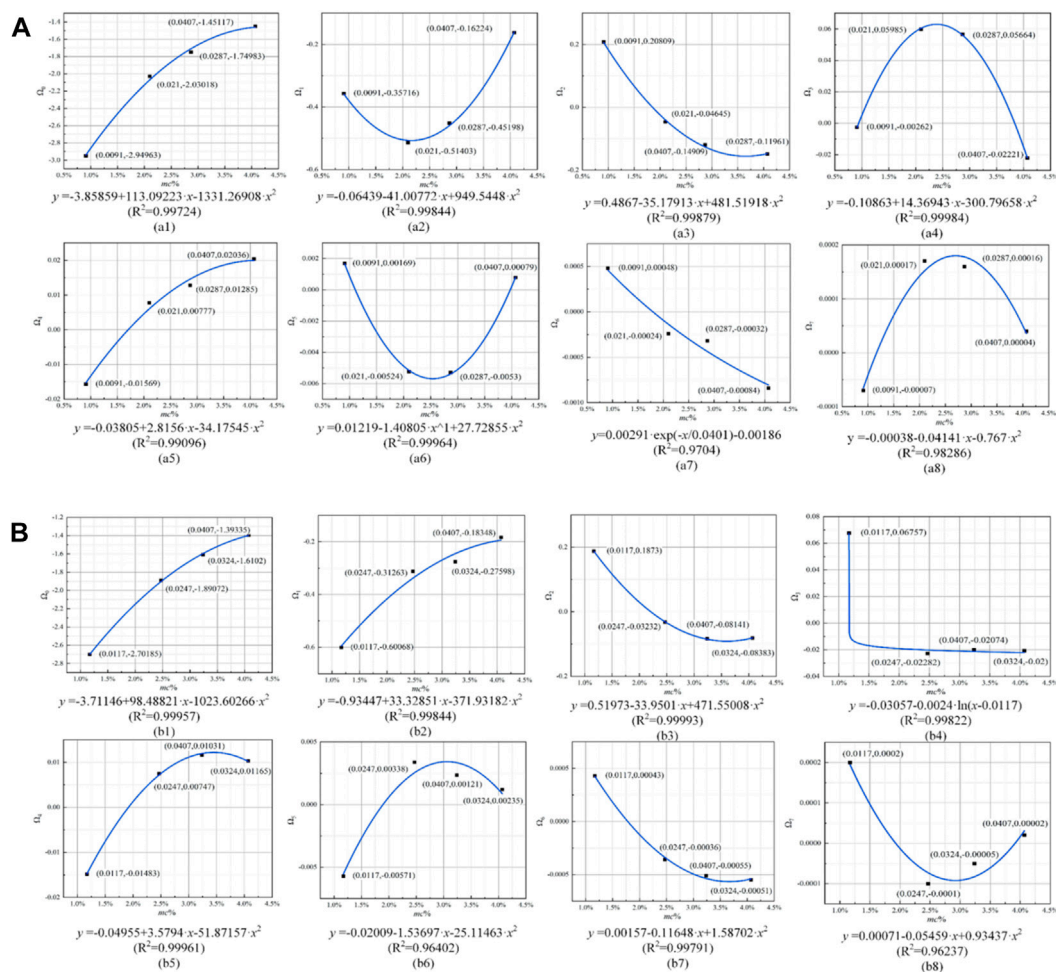


FIGURE 5
(Continued).

corresponding to different values of n ($n = 5, 6, 7, 8$), and then analyze the effect of different values of n on the extraction of the feature parameter in detail.

When $n = 7$ and $n = 8$, it is observed that the measured values of dielectric modulus for the four different degrees of moisture are almost completely overlapped with the corresponding fitted curves, and the R^2 shown in Table 5 are higher than 0.99, which satisfies the fitting requirements. Therefore, through in-depth comparative analysis, this paper selects Taylor's formula at $n = 7$, shown in Eq. 22, to extract the characteristic coefficients of the dielectric modulus.

$$M''(f) = \Omega_0 + \Omega_1 f + \Omega_2 f^2 + \Omega_3 f^3 + \Omega_4 f^4 + \Omega_5 f^5 + \Omega_6 f^6 + \Omega_7 f^7. \quad (22)$$

By determining the values of the coefficients $\Omega_0 \sim \Omega_7$ in Eq. 22, Taylor's formula for the approximate replacement of the imaginary part of the dielectric modulus is obtained. Different moisture and aging states will obtain unique corresponding sets of parameters ($\Omega_0 \sim \Omega_7$), and conversely,

each different set of parameters ($\Omega_0 \sim \Omega_7$) can only characterize unique moisture and aging states. The extracted parameters are shown in Table 6.

Due to the time-consuming preparation cycle of the experiment, there are high requirements for the measurement and preservation of the samples. Therefore, it is difficult to construct a database of dielectric modulus parameters based on laboratory-prepared samples. To overcome the above difficulties, this paper finds that there is a quantitative relationship between the dielectric modulus characteristic parameters $\Omega_0 \sim \Omega_7$ and $mc\%$, as shown in Figure 5. The fitting results show that the fitting degree R^2 of each curve reaches above 0.9, thus effectively establishing a mapping relationship between the moisture content $mc\%$ and the characteristic parameters $\Omega_0 \sim \Omega_7$.

Therefore, based on the curve fitting method, more than 1,000 sets of dielectric modulus parameters of five aging states (unaged, mild aging, moderate aging, severe aging, and insulation life termination) and four moisture states (dry, mild damped, moderate damped, and severe damped) are obtained to construct the parameter database.

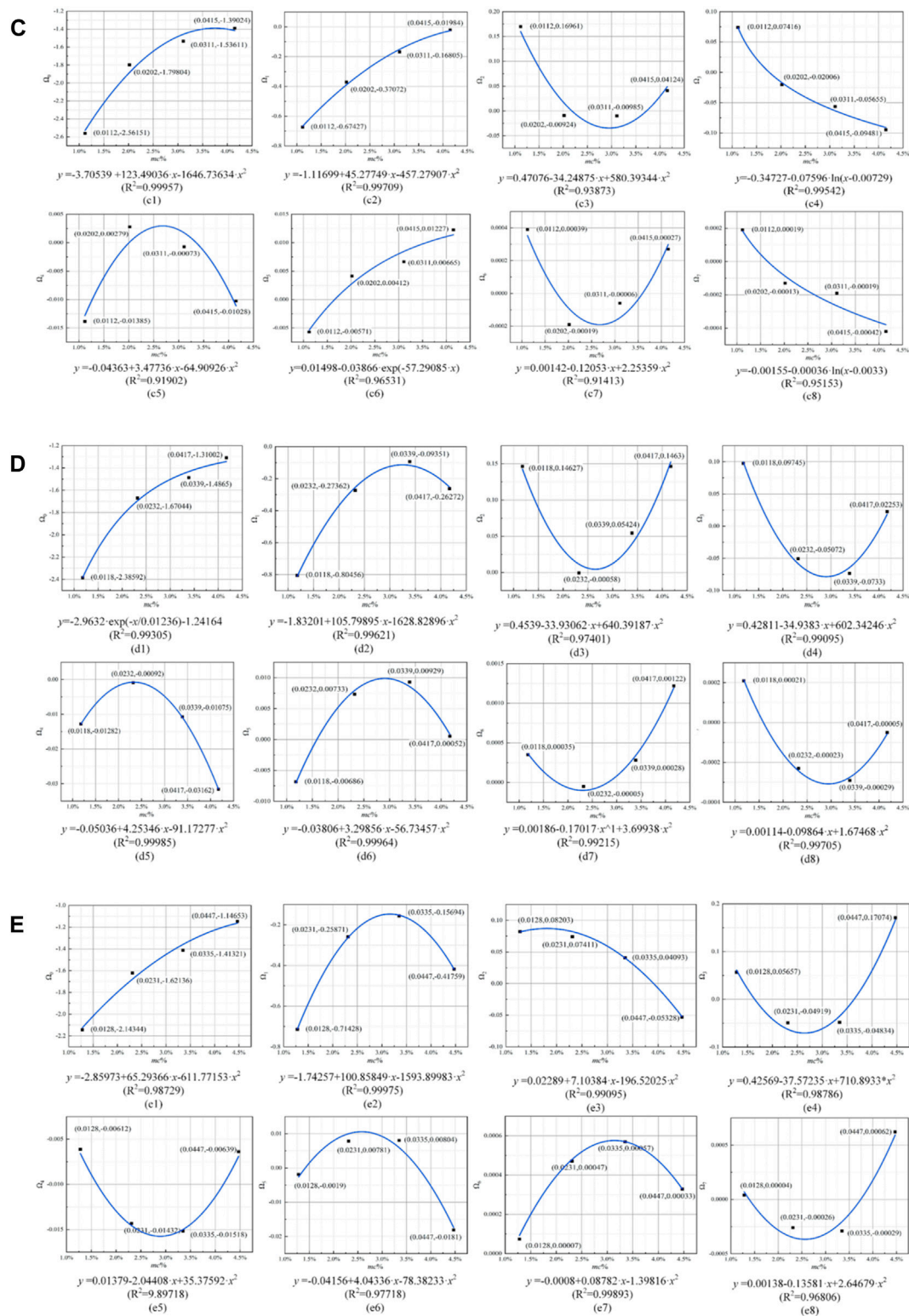


TABLE 7 Details of bushing to be tested.

Number	Voltage level (kV)	Model number	Manufacturer	Moisture (%)	DP	Test temperature (°C)
Bushing 1	110	BRLW-110/630-3	Nanjing Electric Porcelain Factory	$mc\% = 1.27$	731	25
Bushing 2	110	GOB 550SFPS123 107-K	Sweden ABB	$mc\% = 2.38$	684	35
Bushing 3	110	GOB 550SFPS123 107-K	Sweden ABB	$mc\% = 1.87$	267	25

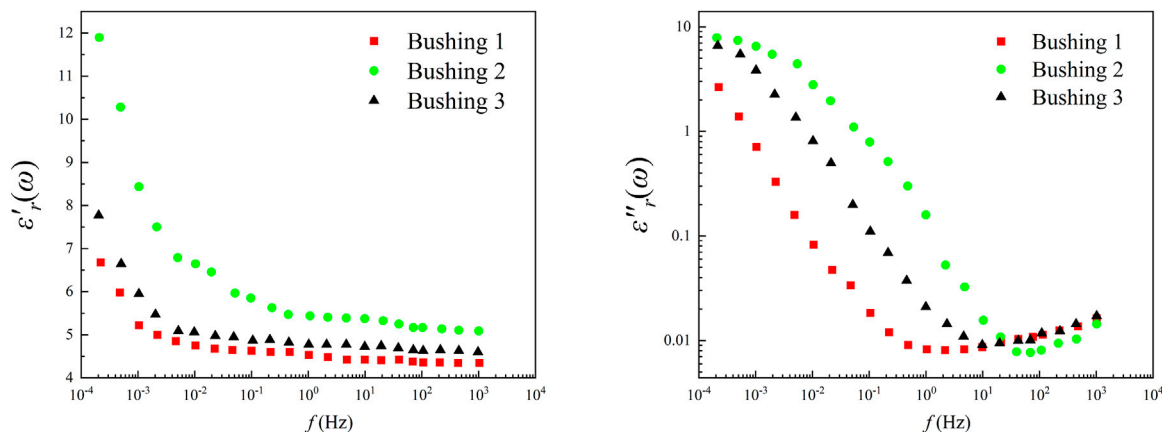


FIGURE 6
The $\varepsilon'_r(\omega)$ and $\varepsilon''_r(\omega)$ curves of bushing to be tested.

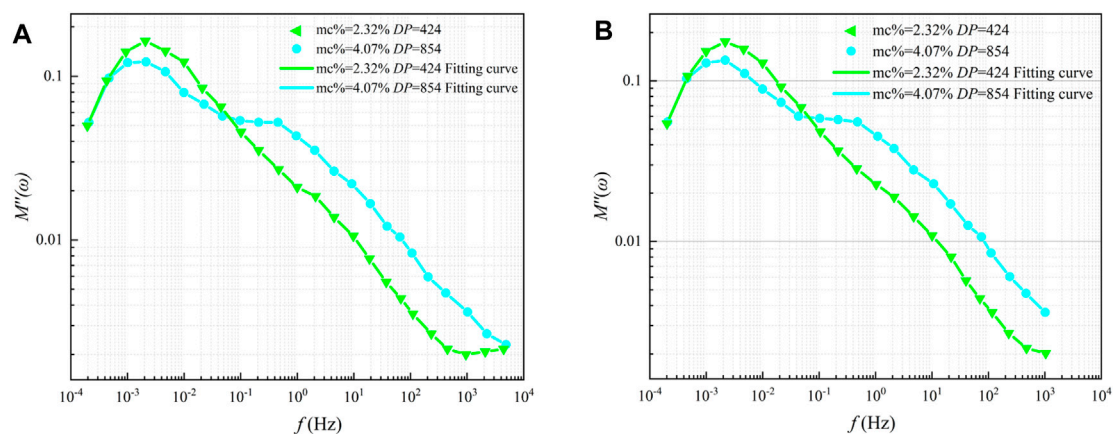


FIGURE 7
(A) $M''(\omega)$ curves with original data; (B) $M''(\omega)$ curves with reducing partial data.

5 Status evaluation results of on-site bushings based on the K-NN algorithm

5.1 Testing and analysis of field bushing

In this paper, three bushings retired from a 110 kV substation of China Southern Power Grid Company are selected for moisture and aging condition assessment to verify the effectiveness and feasibility of the proposed assessment method. The test equipment is a DIRANA dielectric response tester manufactured by OMICRON

company. The Output terminal of DIRANA (high-voltage electrode) is connected to the terminal of the casing; the CH1 terminal (low-voltage electrode) is connected to the measurement terminal of the casing; in order to ensure the safety and reliability of the measurements, it is required that the shell of DIRANA and casing is well grounded.

The test bushing was named bushing 1, bushing 2, and bushing 3. The details of the three bushings are shown in Table 7 and the frequency domain dielectric response test results are shown in Figure 6. After the FDS test, it was disassembled and subjected to

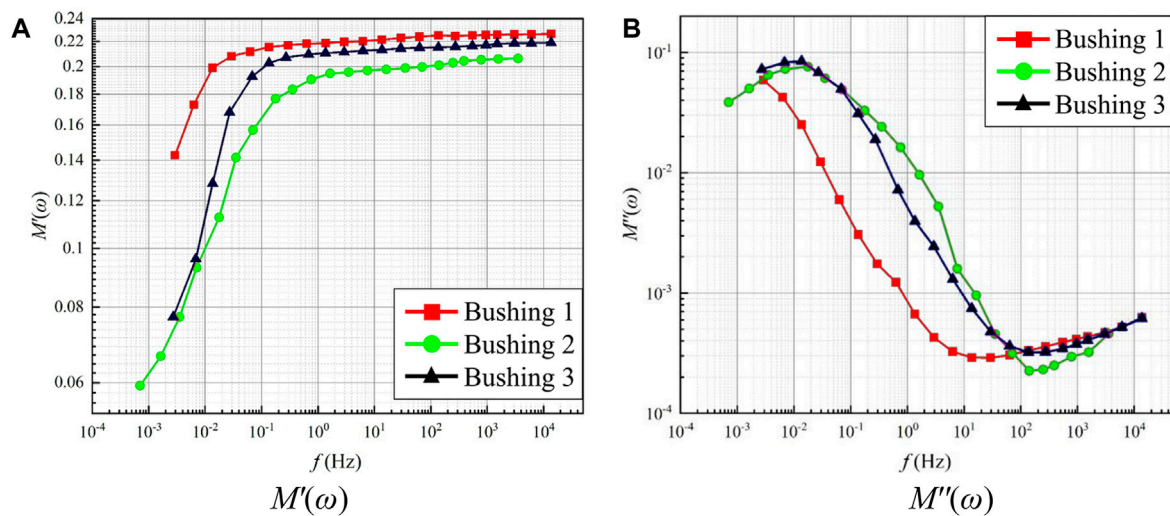


FIGURE 8
FDS curves after temperature correction. (A) $M'(\omega)$. (B) $M''(\omega)$.

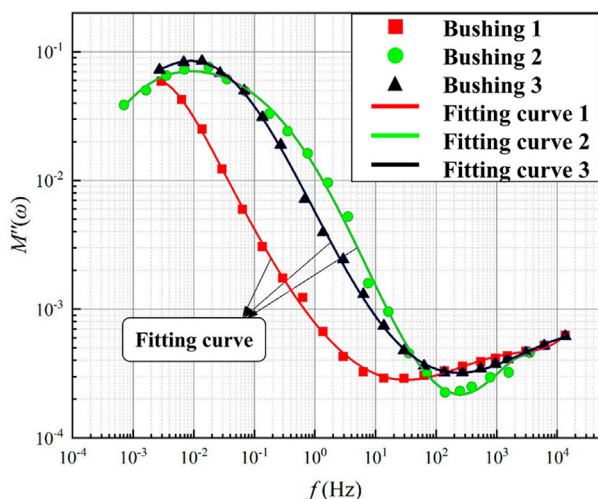


FIGURE 9
Fitting curves of bushing to be tested.

moisture measurements and DP measurements to obtain its true moisture and aging status.

Due to the limitation of the measurement conditions, the frequency range of the field test is $f \in (1 \times 10^{-4} \text{ Hz to } 1 \times 10^3 \text{ Hz})$, as shown in Figure 6. Compared with the FDS test results under

laboratory conditions, the field test is missing part of the test data in the high-frequency region ($1 \times 10^3 \text{ Hz to } 5 \times 10^3 \text{ Hz}$). Therefore, in order to investigate whether the extracted dielectric modulus parameters under the partially missing data condition will generate errors and thus lead to assessment failure, the following discussion is carried out in this paper:

Taking the dielectric modulus $M''(\omega)$ curves for the lightly damped and heavily aged condition ($mc\% = 2.32\%$, $DP = 424$) and the heavily damped and lightly aged condition ($mc\% = 4.07\%$, $DP = 854$) as an example, the $M''(\omega)$ curves for the complete and partially data missing conditions $M''(\omega)$ curves are shown. $M''(\omega)$ curves are shown in Figure 7 below.

5.2 Temperature correction

FDS test results are very sensitive to temperature changes, and changes in the test temperature can result in changes in the test results. Therefore, it is necessary to correct the temperature of the test results of the field bushing. In this paper, the “frequency shift factor” technique is used to obtain the “frequency shift factor” $\alpha(T)$ of each FDS curve by using Eq. 23. Then, by multiplying the measurement frequency by the “frequency shift factor” $\alpha(T)$, as shown in Eq. 24, the sampling frequency at the reference temperature is obtained, and then the FDS curve at the reference temperature is obtained.

TABLE 8 Characteristic parameters extracted from the three bushings.

Number	Characteristic parameter								R^2
	Ω_0	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	
Bushing1	-3.1070	-0.5820	0.1690	0.0220	-0.0010	-0.001	-0.0009	0.0002	0.99912
Bushing2	-1.8964	-0.7753	-0.2018	0.0316	0.0254	0.0034	-0.0011	-0.0002	0.99820
Bushing3	-2.2435	-0.9006	0.01006	0.0886	-0.0055	-0.003	0.00035	0.00001	0.99951

TABLE 9 Evaluation results of moisture and aging state of field bushings.

Number	Predictive status labels	Predictive status labels	Actual parameters (%)	Real state
Bushing1	A ₃ B ₂	Mild moisture	DP = 731, mc% = 1.27	Mild moisture
		Moderate aging		Moderate aging
Bushing 2	A ₃ B ₃	Moderate moisture, Moderate aging	DP = 684, mc% = 2.38	Moderate moisture Moderate aging
Bushing 3	A ₅ B ₂	Mild moisture	DP = 267, mc% = 1.87	Mild moisture
		End of life		End of life

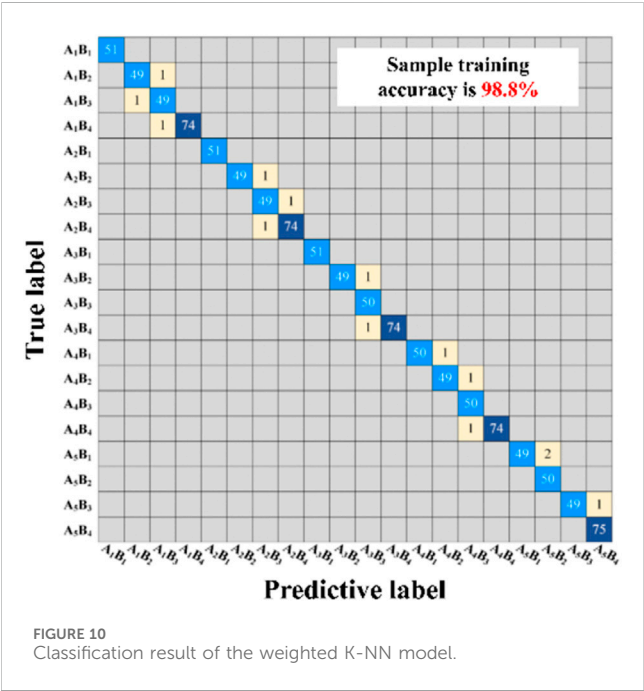


FIGURE 10 Classification result of the weighted K-NN model.

$$\alpha(T) = EXP\left[\frac{E_a\left(1/T_t - 1/T_{ref}\right)}{R}\right], \quad (23)$$

$$f_{ref} = f_t \cdot \alpha(T), \quad (24)$$

where T_{ref} and T_t represent the reference and test temperatures, respectively. f_t and f_{ref} denote the test and reference frequencies, respectively. E_a represents the activation energy of the oil-impregnated cellulose, which is generally considered to be $E_a = 103 \text{ kJ/mol}$ and R is the gas constant, $R = 8.314 \text{ J/mol/K}$.

Since the FDS test was performed in this paper at a constant laboratory ambient temperature of 45°C (318.15 K), the reference temperature $T_{ref} = 318.15 \text{ K}$ in Eq. 23, while the test temperatures of bushing 1, bushing 2, and bushing 3 were 25°C , 35°C , and 25°C ,

respectively. Therefore, the test temperatures in Eq. 23 were sequentially set to 298.15 , 308.15 , and 298.15 K , and the $\alpha(T)$ of bushing 1, bushing 2, and bushing 3 were obtained to be 13.6277 , 3.5384 , and 13.6277 , respectively. Subsequently, based on Eq. 24, the temperature-corrected $\epsilon'(\omega)$ and $\epsilon''(\omega)$ curves and $M'(\omega)$ and $M''(\omega)$ curves, as shown in Figure 8.

After the temperature correction of the field bushing of the unknown state, this paper will use Eq. 22 to extract the characteristic parameters of its dielectric modulus $M''(\omega)$ curves, followed by the moisture and aging state assessment.

5.3 Extraction of dielectric modulus characteristic coefficients of field bushing

Equation 22 is used to extract the characteristic parameters from the temperature-corrected dielectric modulus $M''(\omega)$ curve, and then the characteristic parameters are used to assess the moisture and aging state. As shown in Figure 9, the fitted curves based on Eq. 22 almost coincide with the test data points, and the characteristic parameters Ω_i ($i = 0, 1, \dots, 7$) and the corresponding fit degrees contained in the three bushings are listed in Table 8. The R^2 exceeds 0.998 , which verifies the validity and reliability of the mathematical model of Taylor's formula, and these characteristic parameters can be used to assess the moisture and aging status of the bushings in the field.

5.4 Results of moisture and aging state assessment of field bushings based on the K-NN algorithm

In this paper, 10 samples with the closest distance to the samples to be tested are selected in the training sample set ($K = 10$). Subsequently, the training sample set is established on top of the dielectric modulus feature parameter database, while the field bushing extracts the feature parameter as the samples to be tested, and finally, the K-NN

TABLE 10 Evaluation results of four classification algorithms.

Bushing number	Actual state	K-NN	SVM	NB	DT
1	A ₃ B ₂	A ₃ B ₂	A ₃ B ₂	A ₃ B ₂	A ₃ B ₂
2	A ₃ B ₃	A ₃ B ₃	A ₂ B ₃	A ₄ B ₁	A ₁ B ₄
3	A ₄ B ₂	A ₅ B ₂	A ₅ B ₂	A ₄ B ₂	A ₄ B ₂

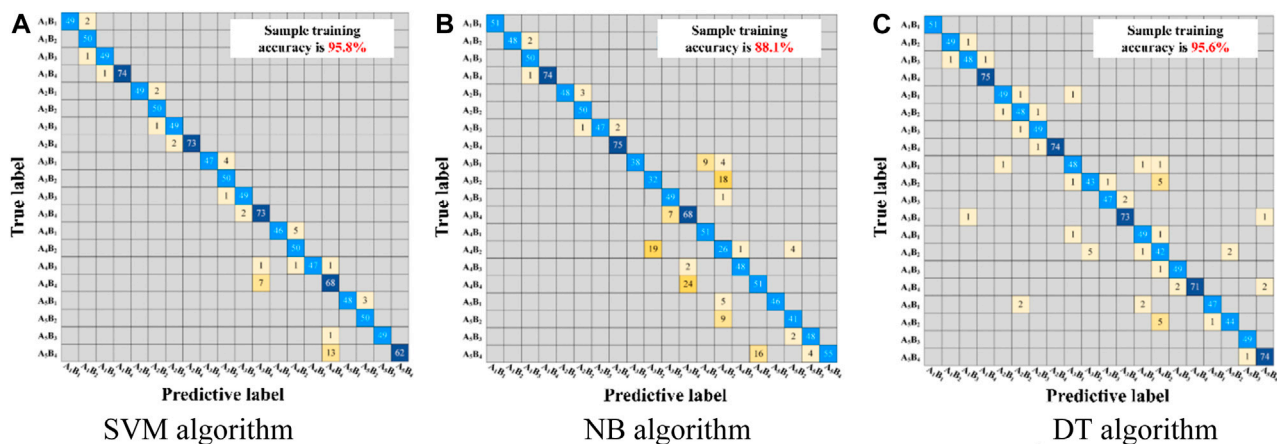


FIGURE 11
Confusion matrix of the other algorithms. (A) SVM algorithm. (B) NB algorithm. (C) DT algorithm.

classification algorithm is applied to evaluate the moisture and aging states of the field bushings. The evaluation results are shown in Table 9 and its confusion matrix is shown in Figure 10.

From the evaluation results of the K-NN classification algorithm obtained in Table 9, it can be seen that the predicted results for bushing 1 and bushing 2 agree with the actual results. The predicted result for bushing 3 is mild moisture and insulation life termination ($DP \leq 250$, $1\% \leq mc\% < 2\%$), however, in fact, the actual moisture content $mc\%$ and DP values of bushing 3 are 1.87% and 267, respectively. and the real state is mild damp and heavy aging. The evaluation method produces an error in the assessment of the aging state of bushing 3. The real aging state of bushing 3 is heavy aging. However, the difference between the real DP value and the assessment result is only 17, and the reasons for the error may be as follows:

Firstly, when the DP measurement of bushing 3 is carried out, the sampling of the oil-impregnated insulation paper of the capacitor core may be concentrated in certain places, which cannot fully reflect the overall aging state of the bushing. Secondly, the DP measurement device is a high-precision instrument, and factors such as changes in the temperature of the testing environment and the quality of the test sample can affect the results of the DP test. Thirdly, it is generally believed that when DP drops to about 200, it no longer undergoes significant changes. Therefore, the insulating paper of the bushing may have reached the threshold value, and DP no longer decreases with the increase of aging time.

In summary, although the method proposed in this paper produces an error in the assessment of the aging state of field bushing 3, it is considered that the error is not caused by the defects of the model itself through the above analysis, and therefore, the prediction method of moisture and aging state of oil-impregnated paper bushing based on the K-NN classification algorithm proposed in this article can be used to effectively assess the bushing in the field and has a high degree of accuracy.

The dielectric modulus characteristic parameter proposed in this article is independent of insulation dimensions, and it is only related to the aging state and moisture content of oil-impregnated

insulation paper. Therefore, the method proposed in this article is applicable to bushings of any voltage level, and the assessment process for on-site bushings is as follows: 1) Measure the internal temperature of the on-site bushing. 2) Use a dielectric response analyzer to perform FDS testing on the on-site bushing to obtain the complex relative permittivity curve at that temperature. 3) Convert the complex relative permittivity curve at the testing temperature to the complex relative permittivity curve at 45°C using the temperature-frequency factor. 4) Convert the complex relative permittivity curve to the dielectric modulus curve using the dielectric modulus model. 5) Extract characteristic parameters from the dielectric modulus curve using the Taylor formula mathematical model. 6) Based on the extracted characteristic parameters, assess the aging and moisture status of the bushing using the K-NN classification algorithm.

5.5 Comparison and validation of evaluation results of each classification algorithm

Other algorithms (Support Vector Machine, Naive Bayes, and Decision Tree) were used to assess the state of the three field bushings respectively. Comparing the assessment results of the four classification algorithms in Table 10, it can be found that the K-NN classification algorithm is the most accurate. the SVM algorithm produces an error in assessing the aging state of bushing 2 and bushing 3. The NB algorithm produces an error in assessing the dampness and aging state of casing 2. The NB algorithm produced errors for both moisture and aging of bushing 2. In addition, the DT algorithm produced large errors in the moisture and aging assessment of bushing 2. The evaluation accuracy was obtained according to Eq. 25.

$$\text{Accuracy} = \frac{\text{Correct evaluation number}}{\text{Total evaluation number}} \times 100\%. \quad (25)$$

Based on Eq. 25, the evaluation accuracy of the four classification algorithms is calculated. Among them, the assessment accuracy of the SVM method, NB method, and DT method is 66.67%, while the

assessment accuracy of the K-NN method is 83.33%, which is higher than the other three algorithms. The above evaluation results verify the effectiveness and feasibility of the evaluation scheme proposed in this paper.

In addition, the confusion matrix [Figure 11](#) shows that the sample training accuracy of the SVM algorithm, NB algorithm, and decision tree algorithm are 95.8%, 88.1%, and 95.6%, respectively, which are lower than that of the K-NN algorithm's sample training accuracy of 98.8%. Therefore, the assessment method of the K-NN algorithm has the highest accuracy.

6 Conclusion

The innovation of this article lies in establishing a Taylor formula mathematical model to extract characteristic parameters for characterizing the aging and moisture degree of insulation paper from the dielectric modulus $M^*(\omega)$ curve of oil-immersed paper bushings and obtaining fitting relationships between various characteristic parameters and insulation degradation status. Then, the KNN classification algorithm was used to evaluate the aging and moisture status of insulation paper for bushings. The main results achieved in the thesis are as follows:

- (1) A mathematical model of dielectric modulus is established, and the effects of moisture and aging on the dielectric modulus curve are analyzed in depth, and it is found that the $M''(\omega)$ curve can better distinguish the effects of aging and moisture, and the low-frequency region of the $M''(\omega)$ curve characterizes the aging state of insulation paper. The aging information characterized by the low-frequency region of the $M''(\omega)$ curve and the moisture information characterized by the high-frequency region of the $M''(\omega)$ curve do not interfere with each other, which proves that the dielectric modulus can be used as a powerful tool for the assessment of the moisture and aging state of the oil-paper insulation.
- (2) The mathematical model of Taylor's formula is established, and it is proved that each coefficient Ω_i ($i = 0, 1, \dots, n$) of Taylor's formula can be used as the characteristic parameter for state evaluation. In addition, the influence of the order n of Taylor's formula on the value of the extracted parameters is investigated, and the results show that a small n will affect the accuracy of the parameters, and a large n will affect the complexity of the calculation. Subsequently, the extraction of the characteristic parameters ($\Omega_0 \sim \Omega_7$) of the dielectric modulus curves was realized on the basis of $n = 7$, and the results showed that the goodness-of-fit R^2 of the characteristic parameters ($\Omega_0 \sim \Omega_7$) to $mc\%$ was more than 0.9, which had high reliability.
- (3) A comprehensive assessment method for the moisture and aging state of bushing oil-paper insulation based on the K-NN algorithm is proposed. By applying this method to assess the moisture and aging state of three field bushings, the results show that the assessment method based on the K-NN

algorithm achieves a training accuracy of 98.8% for the samples, and the accuracy of the assessment results reaches more than 83%, which has an obvious advantage compared with other classification algorithms (Support Vector Machine, Naive Bayes, and Decision Tree).

This research has not only made theoretical contributions but can also serve as a methodological foundation for practical engineering applications.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

ZL: Investigation, Writing–review and editing. YS: Validation, Writing–original draft. HC: Data curation, Writing–original draft. NZ: Data curation, Writing–original draft. BL: Formal Analysis, Writing–review and editing. YL: Validation, Writing–original draft. YF: Investigation, Writing–original draft. YZ: Conceptualization, Writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by North China Power Engineering Co., Ltd. of China Power Engineering Consulting Group under Grant number JBGS 2023-01.

Conflict of interest

Authors ZL, YS, HC, NZ, YL, and YF were employed by North China Power Engineering Co., Ltd. of China Power Engineering Consulting Group.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Akbari, M., Mostafaei, M., and Rezaei-Zare, A. (2023). Estimation of hot-spot heating in OIP transformer bushings due to geomagnetically induced current. *IEEE Trans. Power Deliv.* 38 (2), 1277–1285. doi:10.1109/tpwr.2022.3212322
- Bouaicha, A., Fofana, I., Farzaneh, M., Setayeshmehr, A., Borsi, H., Gockenbach, E., et al. (2009). Dielectric spectroscopy techniques as quality control tool: a feasibility study. *IEEE Electr. Insul. Mag.* 25 (1), 6–14. doi:10.1109/mei.2009.4795464
- Buchacz, J., Cichón, A., and Skubis, J. (2017). Detection of conductive layers short circuit in HV condenser bushings using frequency domain spectroscopy. *IEEE Trans. Dielectr. Electr. Insulation* 24 (1), 552–558. doi:10.1109/tdei.2016.005933
- Choi, K., Singh, S., Kodali, A., Pattipati, K., Sheppard, J., Namburu, S., et al. (2009). Novel classifier fusion approaches for fault diagnosis in automotive systems. *IEEE Trans. Instrum. Meas.* 58 (3), 602–611. doi:10.1109/tim.2008.2004340
- DL, T. (2005). *Guide for the diagnosis of insulation aging in oil-immersed power transformer*.
- Fan, X., Liu, J., Lai, B., Zhang, Y., and Zhang, C. (2021). FDS measurement-based moisture estimation model for transformer oil-paper insulation including the aging effect. *IEEE Trans. Instrum. Meas.* 70, 1–10. doi:10.1109/tim.2021.3070622
- Fofana, I., and Hadjadj, Y. (2016). Electrical-based diagnostic techniques for assessing insulation condition in aged transformers. *Energies* 9 (9), 679. doi:10.3390/en9090679
- Islam, M. M., Lee, G., Hettiwatte, S. N., and Williams, K. (2018). Calculating a health index for power transformers using a subsystem-based GRNN approach. *IEEE Trans. Power Deliv.* 33 (4), 1903–1912. doi:10.1109/tpwr.2017.2770166
- Jadav, R. B., Ekanayake, C., and Saha, T. K. (2014). Understanding the impact of moisture and ageing of transformer insulation on frequency domain spectroscopy. *IEEE Trans. Dielectr. Electr. Insul.* 21 (1), 369–379. doi:10.1109/tdei.2013.003984
- Jyothi, N. S., and Ramu, T. S. (2012). A model for the temperature distribution in resin impregnated paper bushings. *Asia-Pacific Power Energy Eng. Conf. (APPEEC)*, 1–4. doi:10.1109/APPEEC.2012.6307302
- Li, S., Yang, L., Li, S., Zhu, Y., Cui, H., Yan, W., et al. (2020). Effect of AC-voltage harmonics on oil impregnated paper in transformer bushings. *IEEE Trans. Dielectr. Electr. Insulation* 27 (1), 26–32. doi:10.1109/tdei.2019.008247
- Li, Y., Zhang, W., Li, H., Xu, Y., and Zhang, G. (2021). Aging assessment of oil-impregnated-paper electrical equipment via near-infrared spectroscopy powered by improved PCA-RBF-NN: modelling and field practices. *IEEE Trans. Dielectr. Electr. Insulation* 28 (6), 2035–2042. doi:10.1109/tdei.2021.009813
- Liao, W., Zhou, L., Li, Z., Wang, D., Zhang, J., Cai, J., et al. (2021). An IGBA algorithm-based curve reconstruction method of frequency-domain dielectric spectroscopy for OIP bushing with nonuniform moisture distribution. *IEEE Trans. Transp. Electrification* 7 (4), 3194–3203. doi:10.1109/tte.2021.3069014
- Liao, W., Zhou, L., Wang, D., Wang, A., Guo, L., and Cui, Y. (2020). A diagnostic method for moisture intrusion fault in OIP bushing. *IEEE Trans. Instrum. Meas.* 69 (9), 7072–7081. doi:10.1109/tim.2020.2976278
- Linhjell, D., Lundgaard, L., and Gafvert, U. (2007). Dielectric response of mineral oil impregnated cellulose and the impact of aging. *IEEE Trans. Dielectr. Electr. Insulation* 14 (1), 156–169. doi:10.1109/tdei.2007.302884
- Liu, J., Fan, X., Zhang, C., Lai, C. S., Zhang, Y., Zheng, H., et al. (2021). Moisture diagnosis of transformer oil-immersed insulation with intelligent technique and frequency-domain spectroscopy. *IEEE Trans. Industrial Inf.* 17 (7), 4624–4634. doi:10.1109/tii.2020.3014224
- Liu, J., Fan, X., Zhang, Y., Li, S., and Jiao, J. (2020b). Frequency domain spectroscopy prediction of oil-immersed cellulose insulation under diverse temperature and moisture. *IEEE Trans. Dielectr. Electr. Insulation* 27 (6), 1820–1828. doi:10.1109/TDEL.2020.008813
- Liu, J., Fan, X., Zhang, Y., Zheng, H., and Zhu, M. (2020a). Quantitative evaluation for moisture content of cellulose insulation material in paper/oil system based on frequency dielectric modulus technique. *Cellulose* 27, 2343–2356. doi:10.1007/s10570-019-02820-3
- Netaworldjournal, (2018). IEEE guide for dielectric frequency response test. *IEEE Stand. C57* 161, 1–76. doi:10.1109/IEEESTD.2018.8571325
- Poovamma, P. K., Ahmed, T., and Viswanatha, C. (2008). “Evaluation of transformer oil by Frequency domain technique,” in *IEEE international conference on dielectric liquids* (IEEE), 1–4.
- Ren, G., Cheng, J., Zhou, Q., Li, C., and Wu, W. (2019). “The method for aging condition prediction of transformer oil-immersed cellulose insulation based upon the aging kinetic equation,” in *2nd International Conference on Electrical and Electronic Engineering* (EEE 2019), 147–151.
- Rouhafzay, G., and Cretu, A. M. (2020). Object recognition from haptic glance at visually salient locations. *IEEE Trans. Instrum. Meas.* 69 (3), 672–682. doi:10.1109/tim.2019.2905906
- Saha, T. K., and Purkait, P. (2004). Investigation of polarization and depolarization current measurements for the assessment of oil-paper insulation of aged transformers. *IEEE Trans. Dielectr. Electr. Insul.* 11 (1), 144–154. doi:10.1109/tdei.2004.1266329
- Su, Z., Zhang, Y., Du, Y., Su, Y., Deng, J., Goh, H. H., et al. (2023). An enhanced X model of oil-impregnated paper bushings including axial and radial nonuniform thermal aging. *IEEE Trans. Dielectr. Electr. Insulation* 30 (2), 825–833. doi:10.1109/tdei.2022.3228223
- Sumereder, C., and Muhr, M. (2010). Moisture determination and degradation of solid insulation system of power transformers. *IEEE Int. Symposium Electr. Insulation. IEEE*, 1–4. doi:10.1109/ELINSL.2010.5549775
- Teng, C., Ding, Y., Zhang, Y., Huang, M., and Zhan, Z. (2022). Investigation on distribution of electro-thermal coupling fields influenced by HVDC bushing insulation properties. *Front. Energy Res.* 10. doi:10.3389/fenrg.2022.1005470
- Wang, D., Zhou, L., Dai, C., Guo, L., and Liao, W. (2019). Insulation defect diagnostic method for OIP bushing based on multiclass LS-SVM and cuckoo search. *IEEE Trans. Instrum. Meas.* 69 (1), 163–172. doi:10.1109/tim.2019.2895482
- Xie, J., Dong, M., Hu, Y., Zhuang, T., Albarracin-Sanchez, R., and Rodriguez-Serna, J. M. (2019). Modeling oil-paper insulation frequency domain spectroscopy based on its microscopic dielectric processes. *IEEE Trans. Dielectr. Electr. Insul.* 26 (6), 1788–1796. doi:10.1109/tdei.2019.008155
- Yang, L. J., Zou, T., Deng, B., Zhang, H., Mo, Y., and Peng, P. (2019). Assessment of oil-paper insulation aging using frequency domain spectroscopy and moisture equilibrium curves. *IEEE Access* 7, 45670–45678. doi:10.1109/access.2019.2906379
- Zaengl, W. S. (2003). Dielectric spectroscopy in time and frequency domain for HV power equipment, Part I: theoretical considerations. *IEEE Electr. Insul. Mag.* 19 (5), 5–19. doi:10.1109/mei.2003.1238713
- Zhang, M., Liu, J., Jia, H., Chen, Q., Lv, J., and Chen, X. (2019a). Modelling the low-frequency electrode dielectric response based on transformer equivalent oil-paper insulation model. *IET Sci. Meas. Technol.* 13 (5), 700–707. doi:10.1049/iet-smt.2018.5495
- Zhang, M., Liu, J., Yin, M., Jia, H., and Lv, J. (2019b). Assessment on oil-paper insulation aging of transformer based on dielectric response model. *Electr. Power Components Syst.* 47 (13), 1145–1155. doi:10.1080/15325008.2019.1663454
- Zhang, Y., Wei, H., Liao, R., Wang, Y., Yang, L., and Yan, C. (2017). A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed transformers. *J. Electr. Eng. Technol.* 12 (2), 830–839. doi:10.5370/jeet.2017.12.2.830
- Zhou, L., Liao, W., Wang, D., Wang, D., Zhang, G., Cui, Y., et al. (2021). A high-precision diagnosis method for damp status of OIP bushing. *IEEE Trans. Instrum. Meas.* 70, 1–10. doi:10.1109/tim.2020.3047194



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Ruijin Zhu,
Tibet University, China
Fei Mei,
Hohai University, China
Jianyong Zheng,
Southeast University, China

*CORRESPONDENCE

Lei Zhou,
✉ 18845095018@163.com

RECEIVED 27 October 2023

ACCEPTED 02 January 2024

PUBLISHED 12 January 2024

CITATION

Zhou L, Gu S, Liu Y and Zhu C (2024), A novel recognition method for complex power quality disturbances based on Markov transition field and improved densely connected network.
Front. Energy Res. 12:1328994.
doi: 10.3389/fenrg.2024.1328994

COPYRIGHT

© 2024 Zhou, Gu, Liu and Zhu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A novel recognition method for complex power quality disturbances based on Markov transition field and improved densely connected network

Lei Zhou*, Shuifu Gu, Yi Liu and Chaoqun Zhu

State Grid Suzhou Power Supply Company, Suzhou, China

In order to solve the difficulty that complex power quality disturbances (PQDs) are difficult to recognize accurately and efficiently under the new power system background, this paper proposes a novel PQDs recognition method based on markov transition field (MTF) and improved densely connected network (DenseNet). Firstly, the one-dimensional PQDs signal is mapped into the two-dimensional image with clear texture features by using MTF encoding method. Then, a DenseNet-S lightweight network is designed and the convolutional attention module (CBAM) is introduced to improve its feature extraction ability, so as to enhance the performance of the network. Finally, the images are input into the improved model for training and learning, and PQDs recognition is realized through the optimal model. In order to verify the effectiveness of the proposed method, experimental tests are carried out based on IEEE 1159 standard simulation dataset and real-world field measured signals dataset, and compared with existing recognition methods. The results show that the proposed method can effectively improve the recognition accuracy and noise robustness of complex PQDs, and has more advantages in disturbances recognition efficiency. It can meet the recognition accuracy and efficiency requirements of massive and complex PQDs events in engineering applications.

KEYWORDS

power quality disturbances, Markov transition field, densely connected network, convolutional block attention module, disturbances recognition

1 Introduction

With the increase of the penetration rate of distributed source loads, which is mainly dominated by wind and solar power generation and new energy charging piles, the power system presents a typical trend of power electronization (Wang et al., 2021), and the PQDs problem of power grid has shown some new features compared with the past. The typical features mainly include two points: 1) the expansion of the scale of new power system disturbance sources; 2) the coupling and superposition of PQDs are strengthened (Wang and Chen, 2019). The interactive coupling of PQDs will exhibit extremely complex non-stationary fluctuation phenomena and further deteriorate the grid power quality, which will not only bring problems such as sensitive equipment damage, data loss, equipment energy consumption increase and other problems to the end user side, but even lead to large-scale

TABLE 1 Mathematical model of PQDs.

PQDs	Equation	Parameters
Normal	$f(t) = \sin(\omega t)$	$\omega = 2\pi f_0, f_0 = 50\text{Hz}$
Sag	$f(t) = \{1 - \alpha[\varepsilon(t - t_1) - \varepsilon(t - t_2)]\} \sin(\omega t)$	$0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$
Swell	$f(t) = \{1 + \alpha[\varepsilon(t - t_1) - \varepsilon(t - t_2)]\} \sin(\omega t)$	$0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T$
Interruption	$f(t) = \{1 - \alpha[\varepsilon(t - t_1) - \varepsilon(t - t_2)]\} \sin(\omega t)$	$0.9 \leq \alpha \leq 1, T \leq t_2 - t_1 \leq 9T$
Harmonic	$f(t) = \sin(\omega t) + \sum_{i=3,5,7} \alpha_i \sin(i\omega t)$	$0.05 \leq \alpha_i \leq 0.15$
Oscillatory transient	$f(t) = \sin(\omega t) + \alpha e^{-(t-t_1)/\tau} \sin(\beta\omega t) \cdot [\varepsilon(t - t_1) - \varepsilon(t - t_2)]$	$0.1 \leq \alpha \leq 0.8, 0.5T \leq t_2 - t_1 \leq 3T, 0.008 \leq \tau \leq 0.04, 8 \leq \beta \leq 32$
Pulse	$f(t) = \sin(\omega t) + \alpha[\varepsilon(t - t_1) - \varepsilon(t - t_2)]$	$1 \leq \alpha \leq 3, 1\text{ms} \leq t_2 - t_1 \leq 3\text{ms}$
Flicker	$f(t) = [1 + \alpha \sin(\beta\omega t)] \sin(\omega t)$	$0.1 \leq \alpha \leq 0.2, 0.1 \leq \beta \leq 0.5$
Notch	$f(t) = \sin(\omega t) - \text{sgn}(\sin(\omega t)) \times \left\{ \sum_{n=0}^9 k \times [\varepsilon(t - (t_1 + 0.02n)) - \varepsilon(t - (t_2 + 0.02n))] \right\}$	$0.1 \leq k \leq 0.4, 0 \leq t_2, t_1 \leq 0.5T, 0.01T \leq t_2 - t_1 \leq 0.05T$
Spike	$f(t) = \sin(\omega t) + \text{sgn}(\sin(\omega t)) \times \left\{ \sum_{n=0}^9 k \times [\varepsilon(t - (t_1 + 0.02n)) - \varepsilon(t - (t_2 + 0.02n))] \right\}$	$0.1 \leq k \leq 0.4, 0 \leq t_2, t_1 \leq 0.5T, 0.01T \leq t_2 - t_1 \leq 0.05T$

power outage accidents in serious cases (Cui et al., 2022). PQDs recognition algorithm is mainly used in the monitoring and management of power systems and power equipment, which can help us detect and identify power quality problems, such as voltage sag, harmonic, flicker, etc., which is crucial to ensure the stable operation of power system, prevent equipment damage and improve the electrical energy use efficiency. Therefore, accurate and efficient classification and recognition of PQDs under the new power system background is the basic requirement to ensure the safe, stable and economic operation of power grid.

The traditional PQDs recognition methods are based on manually extracting features, constructing feature matrices, and using the feature data to train classification model to realize disturbances recognition. Amongst these, feature extraction mostly applies signal processing techniques such as Fourier transform (Huang et al., 2016), Stockwell transform (Yin et al., 2021; Cui et al., 2022), and wavelet transform (Wu et al., 2022). After manual feature extraction, classification algorithms such as support vector machine (Tang et al., 2020), decision tree (Huang et al., 2017a), and artificial neural network (Li et al., 2020) are combined to establish the mapping relationship from continuous features to discrete labels to realize the classification and identification of PQDs. However, the parameter selection and processing process of the above methods are cumbersome, and heavily rely on expert experience, which makes it less generalizable. With the growing penetration of new energy sources, the complexity of PQDs increases accordingly, and the classification methods based on artificial feature extraction are difficult to meet the needs of PQDs classification under the new power system background.

In recent years, deep learning technology has become an important research hotspot in the field of PQDs recognition due to its excellent generalization performance and feature self-extraction capability. Among them, the deep learning methods that have been widely researched and applied are the

Convolutional Neural Network (CNN), which can extract spatial features, and the Recurrent Neural Network (RNN), which has the memory ability of temporal features, to mine the potential relationship between spatial and temporal features of one-dimensional PQDs signals, and then recognize the specific type of PQDs (Ahmadi and Tani, 2019; Wang and Chen, 2019; Sindi et al., 2021). However, although the one-dimensional convolutional layer used in this method can extract temporal features to a certain extent, its temporal characteristics and disturbances classification ability are significantly reduced, and the problem of network gradient vanishing is serious. At the same time, it is not effective in dealing with the problem of feature extraction for multi-type coupling PQDs. To overcome the above problems, some scholars combine the advantages of machine vision and propose the visual conversion of one-dimensional signals, such as Gramian Angular Field (GAF), combined with the current mainstream image classification networks, such as CNN, deep residual network (ResNet), to discover the absolute temporal and spatial relationships of the original signal, and then extract more complete deep-level features of disturbance signals. He et al. (2023) and Jyoti et al. (2021) use the GAF color coding method to convert one-dimensional disturbance signals into rectangular images with different pattern features, which are used as inputs to CNN and ResNet for training and classification respectively, making up for the defects of artificial feature selection. However, GAF uses the Cartesian coordinate system to encode sequences in polar coordinates, and then converts them into Gram matrix using trigonometric operations, which is a complex cross-domain encoding conversion process, computationally intensive, and prone to aliasing with noise. At the same time, its conversion efficiency is on the low side, which makes it difficult to satisfy the recognition efficiency demand of massive and complex PQDs events in engineering practice.

Overall, there have been limited investigations into integrating signal visualization with image classification networks to recognize complex PQDs. Existing studies mainly concentrate on enhancing the disturbances recognition accuracy of the overall method, ignoring the performance of visualization imaging technology and deep learning model in terms of conversion and recognition efficiency. With the massive increase of power quality monitoring data, the recognition efficiency of PQDs is becoming increasingly crucial for power system health condition monitoring. To address this challenge, a novel PQDs recognition method based on MTF and improved DenseNet-SC is proposed in this paper. The proposed method utilizes the MTF encoding method with straightforward and efficient conversion process and clear features to color code the original disturbance signal to form PQDs feature image, and then designing lightweight network DenseNet-S and integrating CBAM attention mechanism to improve the feature capturing ability of the model. Finally, based on IEEE standard and the real-world field data collected by the substation, simulation and measurement dataset of PQDs are established to test the proposed method. The experimental results verify the effectiveness and superiority of the proposed method in PQDs recognition efficiency and precision.

2 PQDs data visualization based on MTF

2.1 PQDs model construction

For various complicated PQDs problems in power system, the real data is difficult to be collected by equipment. Therefore, according to IEEE Std, 2019 power quality standard, mathematical modeling of PQDs signal is carried out in this paper. The nine categories for common basic disturbance signals are sag, swell, interruption, harmonic, oscillatory transient, pulse, flicker, gap, and spike. Corresponding mathematical models for these categories are presented in Table 1.

Complex PQDs are typically overlaid with multiple basic disturbances, including a variety of different categories and different start and end times. The resulting composite waveforms exhibit a complex, cumbersome, and irregular pattern, posing challenges in accurately recognizing the disturbance types.

2.2 Markov transition field

The conversion methods of one-dimensional PQD signal into two-dimensional image are currently mainly investigated using the GAF and its variants. However, due to the need for cross-domain encoding and matrix operation of sequence data, the conversion process of GAF is complex and low efficiency. In addition, the loss of time sequence feature information is easy to occur when processing disturbance signals in noisy environment.

To address the above shortcomings and deficiencies, this paper proposes a visual conversion method of PQDs based on MTF. The correlation between amplitude and time is the main relationship in time-series data. MTF can convert one-dimensional time series data into two-dimensional feature images by considering the time and position information on

the basis of Markov chain and using Markov transition probability for coding, so as to maintain the time order and statistical dynamics in the generated images (Wang and Oates, 2015). Due to its excellent time-series information retention ability, MTF coding technology has been partially applied in fault diagnosis (Yan et al., 2022) and surface electromyography signal analysis (Li et al., 2022). However, as far as we know, MTF coding technology has not been applied to PQDs recognition research in the existing literature.

Given a set of time series signal $X = \{x_1, x_2, \dots, x_N\}$, x_i is the i -th sampling signal point. MTF method firstly divides the time series signal into Q quantile units q_j ($j \in [1, Q]$) according to its amplitude for discretization processing, quantizes each value, and any numerical point x_i can be mapped to the corresponding quantile q_i . Then, by calculating the probability of sampling signal point x_{t-1} and x_t transferring from region q_i to q_j in the form of a first-order Markov chain along the time axis, and taking each calculated probability p as element w_{ij} , the Markov state transition matrix W of $Q \times Q$ dimension is constructed. The expression is shown in Eq. 1:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1Q} \\ w_{21} & w_{22} & \cdots & w_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ w_{Q1} & w_{Q2} & \cdots & w_{QQ} \end{bmatrix} = \begin{bmatrix} p_{1,1}(x_t \in q_1 | x_{t-1} \in q_1) & \cdots & p_{1,Q}(x_t \in q_1 | x_{t-1} \in q_Q) \\ p_{2,1}(x_t \in q_2 | x_{t-1} \in q_1) & \cdots & p_{2,Q}(x_t \in q_2 | x_{t-1} \in q_Q) \\ \vdots & \ddots & \vdots \\ p_{Q,1}(x_t \in q_Q | x_{t-1} \in q_1) & \cdots & p_{Q,Q}(x_t \in q_Q | x_{t-1} \in q_Q) \end{bmatrix} \quad (1)$$

Nonetheless, Markov chain is memoryless, and the probability of state transition at the current moment only depends on the state at the previous moment, without considering the dynamic probability transition of time series data. Therefore, the Markov state transition matrix constructed by Markov chain is also memoryless, which completely ignores the dependence of time step on one-dimensional time series signal X . If the Markov state transition matrix is used directly, a large amount of one-dimensional time series information will be lost. To solve this problem, MTF is used to improve it. By considering the time and position relationship, each transition probability is arranged in time order, and the Markov state transition matrix W is extended to the MTF matrix M across time scales, preserving the time correlation of the original signal. The expression of M is shown in Eq. 2:

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1N} \\ M_{21} & M_{22} & \cdots & M_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ M_{N1} & M_{N2} & \cdots & M_{NN} \end{bmatrix} = \begin{bmatrix} p_{1,j}(x_1 \in q_i | x_1 \in q_j) & \cdots & p_{1,j}(x_1 \in q_i | x_N \in q_j) \\ p_{1,j}(x_2 \in q_i | x_1 \in q_j) & \cdots & p_{1,j}(x_2 \in q_i | x_N \in q_j) \\ \vdots & \ddots & \vdots \\ p_{i,j}(x_N \in q_i | x_1 \in q_j) & \cdots & p_{i,j}(x_N \in q_i | x_N \in q_j) \end{bmatrix} \quad (2)$$

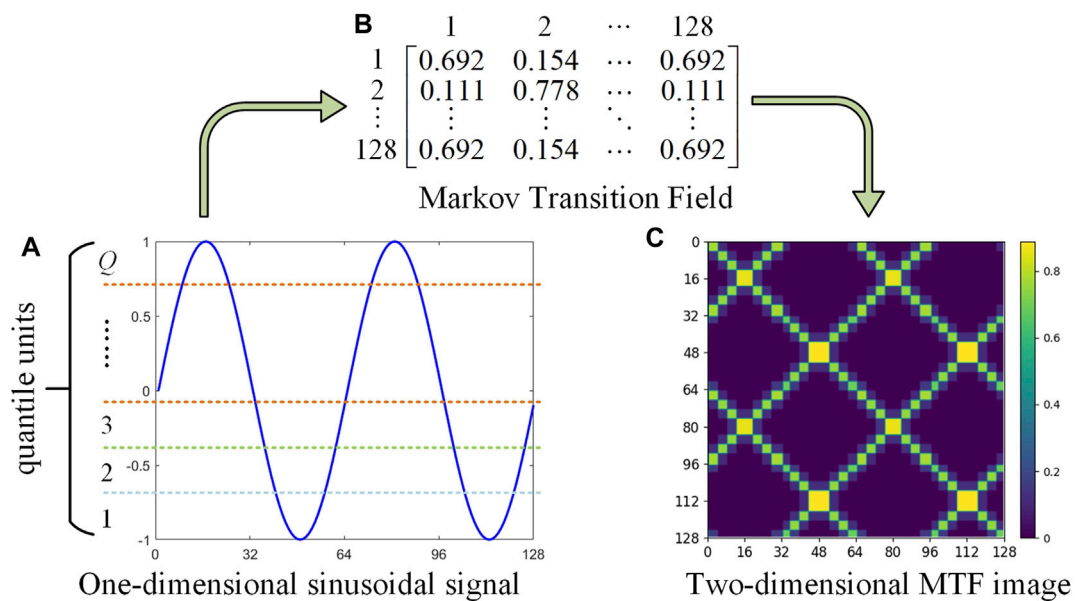


FIGURE 1
MTF visualization image generation process.

where M_{mm} represents the probability of transferring from quantile region q_i corresponding to sampling signal point x_m to region q_j corresponding to x_n , that is, the transition probability of the quantile relationship between q_i and q_j on the matrix W , while the elements on the diagonal are the corresponding self-transfer probability.

Taking one-dimensional sinusoidal signal as an instance, according to the above MTF dynamic transfer information encoding method, its two-dimensional MTF visual image generation process is shown in Figure 1.

Considering that when N is large, using the original MTF to directly generate images will make the images too large and occupy more computer storage space, which is not conducive to the rapid calculation and analysis of on-site intelligent devices or grid background systems. For this purpose, fuzzy kernel $\{1/n^2\}_{n \times n}$ is used to average each non-overlapping pixel to obtain a two-dimensional MTF image of the aggregated $n \times n$ dimension, that is, an aggregated image with dynamic transition probability, which is used as the image modal input data for power quality analysis.

The MTF image coding method offers a technique for visualizing sequences that maintains time dependency. The use of MTF to convert time series signals has the following advantages:

- (1) By considering the dependence between each quantile unit and time step, the correlation between moments can be effectively represented, and the time series information loss of one-dimensional sequence signals can be avoided;
- (2) The one-dimensional time-series signal and MTF image coding method are mapped relations, preventing the loss of feature information.
- (3) The pixel amplitude information is the value of M_{ij} , and the color depth reflects the transformation probability from q_i to

q_j . The two-dimensional image features after visualization are clear and easy to distinguish, which is very conducive to the learning and recognition of deep learning network.

- (4) Compared with GAF, the MTF conversion process is concise and computationally efficient. The introduction of quantile division makes MTF more resistant to interference and noise.

2.3 Visualization of PQDs signal

To achieve a two-dimensional visualization of PQDs data while preserving the temporal correlation and full feature details of signals, this study utilizes the MTF coding technique to convert one-dimensional PQDs signal into two-dimensional MTF image. Based on the mathematical model of PQDs presented in Table 1 and the MTF conversion process detailed in Section 2.2, the fundamental frequency is set to 50 Hz, with the sampling frequency of 3.2 kHz and the sampling length of 10 cycles. Consequently, MTF feature images are generated corresponding to the nine basic PQDs signals as illustrated in Figure 2. The size of the images' horizontal and vertical axes represents their respective dimensions. The conversion result order is in line with the PQD types' order in Table 1.

As depicted in Figure 2, the MTF feature image differentiates time-series signals of various PQDs types at the same spatial location by employing dissimilarities in pixel colors and texture shapes. The feature information of each image is lucid and easily distinguishable, thereby facilitating feature extraction after disturbances composite superposition, and laying a foundation for deep learning network to accurately recognize PQDs.

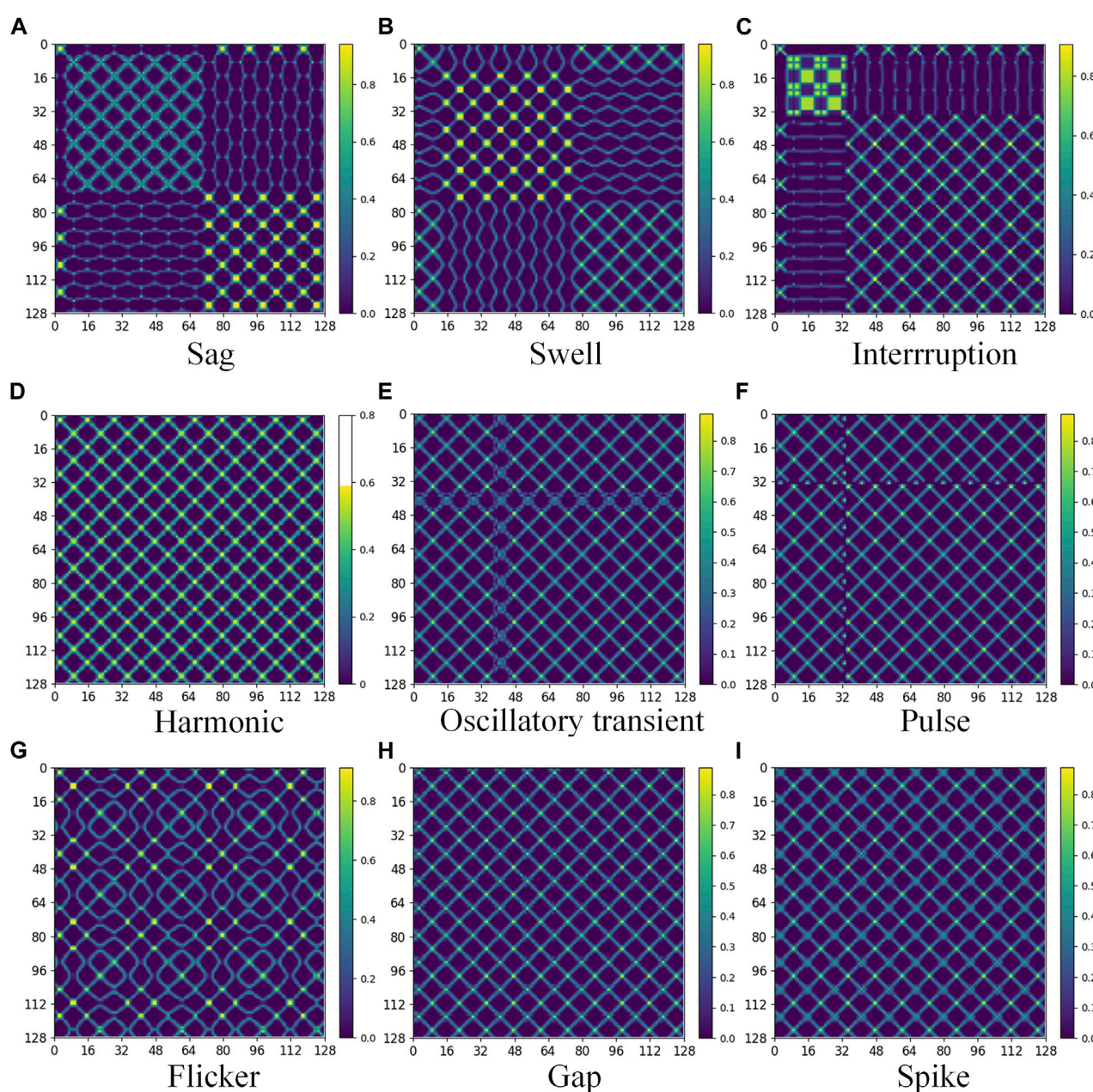


FIGURE 2
MTF visualization image of PQDs basic signals.

3 PQDs recognition based on improved DenseNet

3.1 Fundamental principle of DenseNet

Traditional CNNs come with an abundance of parameters and frequently encounter two primary issues: first, overfitting is easy to occur owing to limited training data, and second, the network layers tend to be shallow, causing inadequate extraction of more advanced information. Deeper networks produce more distinguishable characteristics by acquiring superior-level feature maps, thus it's easier for them to recognize inherent and underlying features.

However, deep neural networks often encounter issues such as gradient vanishing and explosion, which can negatively impact network training and recognition performance (Huang et al., 2017b).

DenseNet incorporates a bypass connection approach similar to ResNet and establishes a dense connection mechanism between convolutional layers, enhancing feature reuse through inter-channel splicing (Huang et al., 2017a). As a result, DenseNet effectively resolves the aforementioned issues. As its core component module, the dense block (DB) structure as depicted in Figure 3, which is designed to ensure maximum information flow between network layers, where: x_0 is the input

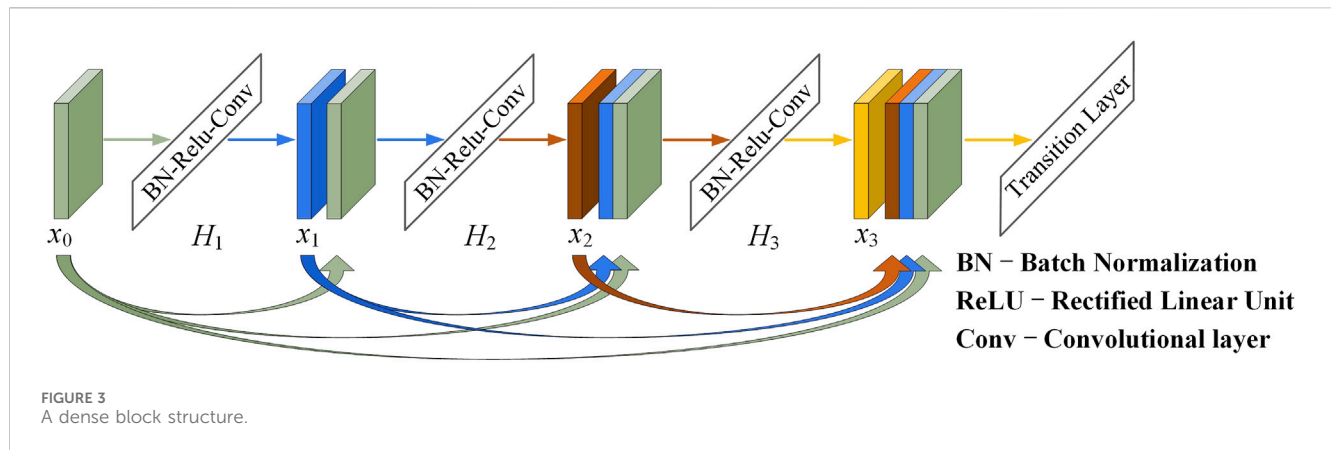


TABLE 2 Network structure design of DenseNet-S.

Network layer	DenseNet-S ($k = 48$)	Output size (input size: $128 \times 128 \times 3$)
Convolution	7×7 Conv, stride 2	$64 \times 64 \times 64$
Pooling	3×3 Max pool, stride 2	$32 \times 32 \times 64$
Dense Block 1	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix} \times 3$	$32 \times 32 \times 208$
Transition Layer	1×1 Conv	$16 \times 16 \times 104$
	2×2 Average pool, stride 2	
Dense Block 2	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix} \times 6$	$16 \times 16 \times 392$
Transition Layer	1×1 Conv	$8 \times 8 \times 196$
	2×2 Average pool, stride 2	
Dense Block 3	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix} \times 9$	$8 \times 8 \times 628$
Transition Layer	1×1 Conv	$4 \times 4 \times 314$
	2×2 Average pool, stride 2	
Dense Block4	$\begin{bmatrix} 1 \times 1 \text{ Conv} \\ 3 \times 3 \text{ Conv} \end{bmatrix} \times 6$	$4 \times 4 \times 602$
Classification Layer	Adaptive average pool	$1 \times 1 \times 602$
	1000D fully-connected, softmax	

feature information, the input of H_1 is x_0 , while the inputs of H_2 are x_0 and x_1 , where x_1 is the output of H_1 , and so on. In this architecture, each layer uses inputs from all previous layers and passes its corresponding feature maps to all subsequent layers. Therefore, this network structure can extract more global and important features, which can achieve more accurate and efficient training effect.

Unlike the conventional convolutional network structure, the number of connections increases to $L(L+1)/2$ from L when the number of DB layers is L . As a result, the feature map obtains from the L -th layer is the outcome of the feature maps that are produced from all previous layers after being spliced in the channel dimension using the following Eq. 3:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (3)$$

where x_l represents the output of layer l . The splicing operation of the output feature map of each layer is denoted by $[x_0, x_1, \dots, x_{l-1}]$. $H_l(\cdot)$ represents a nonlinear combinatorial function, which consists of a batch normalization layer (BN), a linear activation function (ReLU), and a 3×3 convolutional layer (Conv).

Each DB includes multiple convolutional layer structures with the same padding utilized for splicing operations. Although this structure adopts a densely connected pattern, it requires fewer parameters than traditional CNNs. In fact, this network architecture eliminates the need to learn redundant information, reduces the number of feature maps required at the network layer,

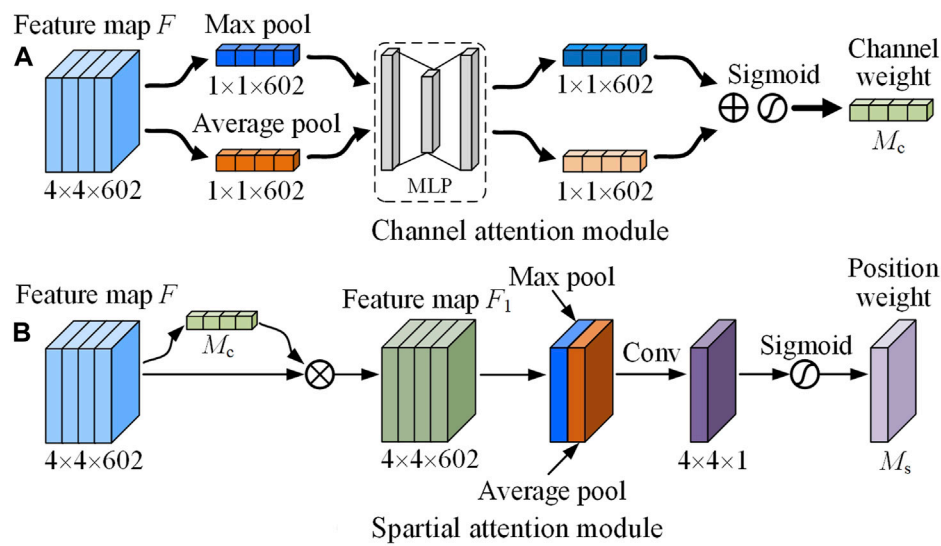


FIGURE 4
Schematic diagram of CBAM.

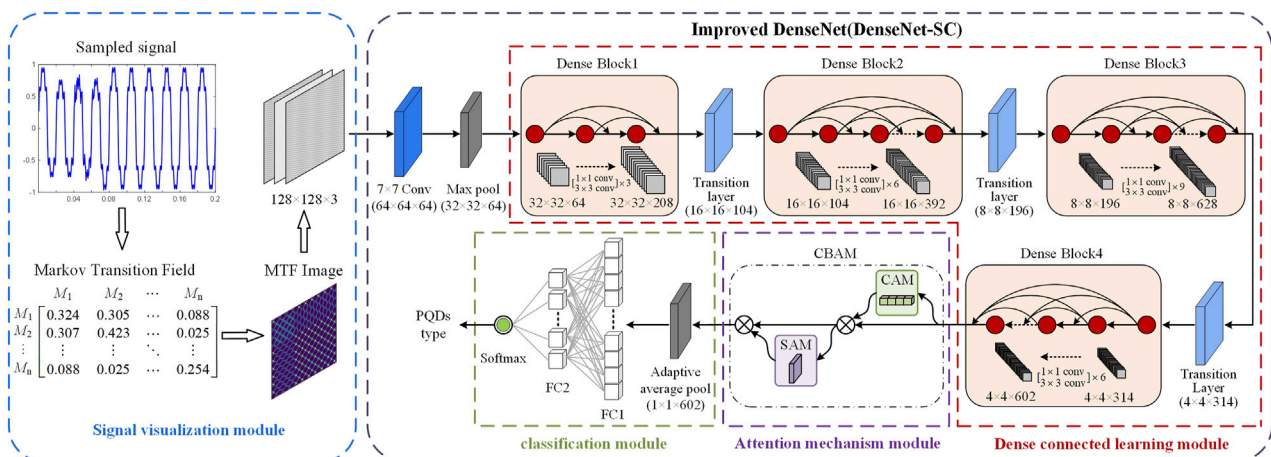


FIGURE 5
PQDs recognition framework based on MTF and improved DenseNet.

and significantly improves parametric efficiency. On the other hand, the continuous concatenation of different layers requires each layer to access the gradient from the original input data and the loss function. This fast access improves the information flow between layers, mitigates the gradient vanishing issue, and facilitates the extraction of deeper semantic information.

3.2 Foundational network architecture design

To avoid overfitting and create a more lightweight network, this paper designs a more compact and lightweight architecture termed DenseNet-S through numerous experimental tests. This architecture serves as the backbone of deep learning network

while ensuring the network's classification precision. Its network structure is shown in Table 2. Among them, the DenseNet-S network contains four groups of DBs, each group is composed of 3, 6, 9, and 6 groups of convolutional layer superposition connections, and the growth rate $k=48$. To enhance the network's computational efficiency and compactness, bottleneck layer and transition layer are introduced for feature map dimensionality reduction. The bottleneck layer is located in DB and consists of 1×1 convolutional layers preceding 3×3 convolutional layers. The transition layer is located between two sets of DBs and consists of a batch normalization layer, a 1×1 convolutional layer, and a 2×2 average pooling layer. For m feature maps output by DB, θ times will be reduced after conversion by transition layer. $\theta \in (0, 1)$ is the compression coefficient. In this paper, we use $\theta = 0.5$.

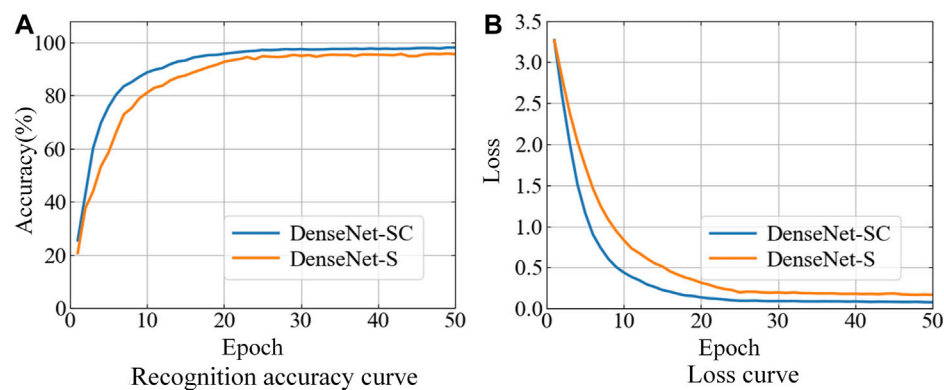


FIGURE 6
Comparison of recognition results on validation set.

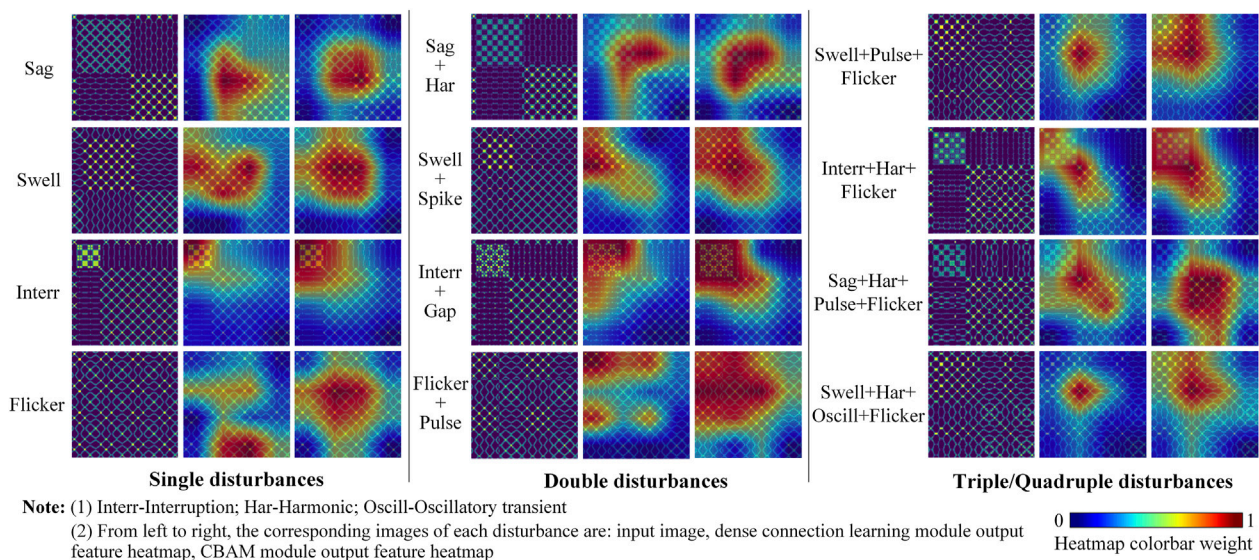


FIGURE 7
Comparison of the class activation heatmap output before and after the CBAM module of improved model.

3.3 CBAM attention mechanism

Accurate extraction of feature image information is crucial for improving the recognition accuracy of PQDs. The convolutional and pooling operation of CNNs defaults the importance of each channel in the feature map to be the same, but due to the different importance of information carried, it is unreasonable to identify the same importance of channels. Based on the processing mechanism of the human visual system, CBAM (Woo et al., 2018) performs dynamic weighted processing of features through autonomous learning on spatial domains and feature channels to enhance the capture of key feature information of images while reducing the interference of non-key information, thus improving the recognition precision. Therefore, this paper introduces CBAM attention mechanism to improve the ability of feature extraction

network to focus details in MTF images corresponding to various PQDs, so as to obtain better recognition effect.

CBAM comprises two modules: channel attention module (CAM) and spatial attention module (SAM). CAM assigns weight coefficients to the feature channels based on their significance. Firstly, all channels information is aggregated through average-pooling and max-pooling to generate two different one-dimensional feature vectors, and then, after fully connected layer operation, element by element addition is performed, channel attention weight M_c is generated through activation operation, which is then multiplied element by element with the input feature map to obtain the channel attention weighted feature map F_1 . The calculation process of CAM module is shown in Figure 4A, and its calculation expression is shown in Eq. 4:

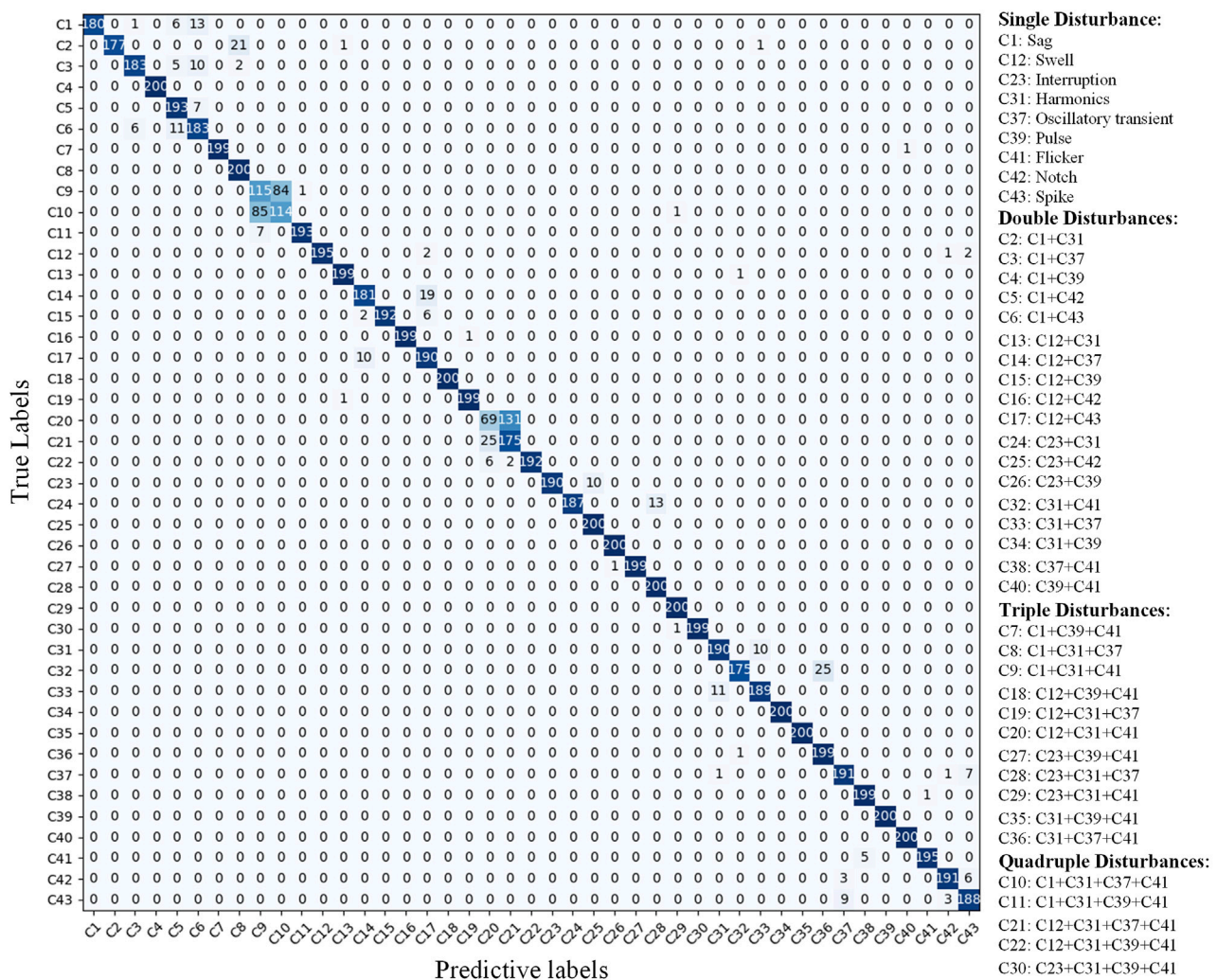


FIGURE 8
Confusion matrix of 43 types PQDs recognition results (SNR = 20 dB).

$$M_C(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (4)$$

where F represents the input feature map, AvgPool and MaxPool indicate average-pooling and max-pooling operations, respectively. MLP denotes the fully connected operation, while σ represents the sigmoid function.

The SAM module uses the feature map F_1 as the input, which has been weighted by CAM. As shown in Figure 4B, the input feature map F_1 undergoes spatially max-pooling and average-pooling resulting in two two-dimensional feature maps, which are splicing for convolution operation and activation operation, and finally generates spatial attention weight M_s to obtain features weighted by SAM. The formula for calculating the spatial attention is shown in Eq. 5:

$$M_s(F) = \sigma(f^{3 \times 3}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f^{3 \times 3}([F_{\text{avg}}^S; F_{\text{max}}^S])) \quad (5)$$

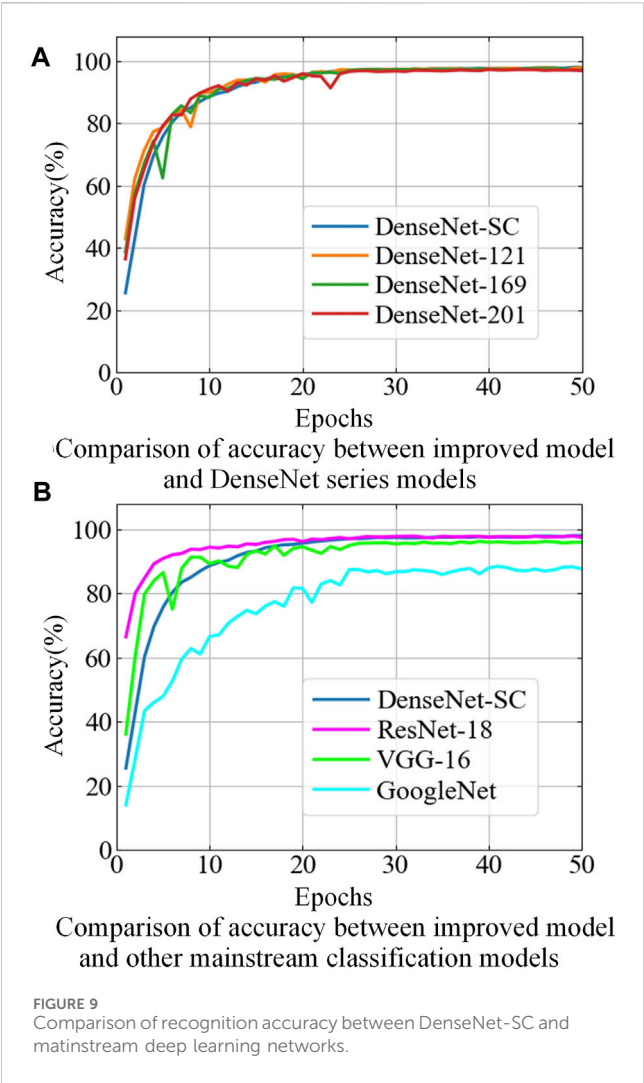
where $f^{3 \times 3}$ represents the dimensionality reduction convolution operation with a convolution kernel size of 3×3 . (The study discovers that using 3×3 convolution kernel size for

dimensionality reduction of feature channel results in better intelligent network recognition performance than 7×7 convolution kernel size. Therefore, the convolution operation with the kernel size of 3×3 is utilized here).

3.4 PQDs recognition framework construction

Combining the imaging advantages of MTF with DenseNet's efficient training and deep feature extraction capabilities, this paper proposes a novel PQDs recognition method based on MTF and improved DenseNet. The PQDs recognition framework is shown in Figure 5. The whole framework mainly includes four parts: signal visualization module, dense connection learning module, attention mechanism module and classification module.

- (1) Signal visualization module: The one-dimensional PQDs time-series sampling signals are deconstructed through MTF dynamic coding, reorganized into MTF matrix to



retain the original temporal information. Finally, the matrix is mapped into a two-dimensional image with easy-to-recognize features and temporal correlation, having a pixel size of 128×128 . This approach indirectly enhances the recognition precision of disturbance types from the data level.

- (2) Dense connection learning module: This module comprises DBs and transition layers. RGB feature maps undergo convolution and dimension reduction via 7×7 convolution layers and 3×3 max-pooling layers to extract shallow feature information. Following feature extraction, the feature information needs to traverse four sets of DBs, with a transition layer connecting every two sets of DBs. When passing through DBs, feature reuse is performed on feature maps from different layers in channel dimension based on dense connection mechanism, which is conducive to extracting deeper feature information and combining shallow layer and deep layer information in the network, which greatly improves feature utilization. Simultaneously, transition layers are introduced to reduce feature dimensions, which reduces network parameter redundancy and enhances the overall efficiency of the model's learning and computation.
- (3) Attention mechanism module: Adding the CBAM attention module between the dense connection learning module and the classification module of DenseNet-S can effectively enhance the network's recognition precision, as evidenced by extensive experimentation. The feature map F output by Dense Block 4 enters the CAM module, where max-pooling and average-pooling operations are performed and result in two $1 \times 1 \times 602$ one-dimensional feature vectors. Then the two feature vectors are sent into the full connection layer for calculation and sum operation. After activation operation, the CAM module generates channel attention weights M_c , which are then multiplied with the feature map F to derive F_1 that is the input feature map for the SAM module. In the SAM module, the feature map F_1 undergoes separate max-pooling and average-pooling operations according to spatial position, and the two results are splicing to generate a $4 \times 4 \times 2$ feature map. Subsequently, this map undergoes dimensional reduction through a 1-channel convolution layer. The spatial attention weight M_s is then obtained by Sigmoid function activation. Ultimately, the feature obtained by multiplying the weights M_s with the input feature map F_1 represents the feature enhanced by CBAM.

TABLE 3 Comparison of recognition performance on Test set.

Recognition network	Recognition accuracy/%				Computational efficiency metrics			
	0 dB	50 dB	30 dB	20 dB	Params/ 10^6	FLOPs/ 10^9	Model Size/MB	Time/ms
GoogleNet	88.09	87.35	86.23	81.24	24.46	5.75	93.47	92.84
VGG-16	96.28	95.67	94.37	90.79	134.44	15.47	512.84	65.31
ResNet-18	97.85	97.23	96.43	92.91	11.20	1.82	42.76	43.05
DenseNet-121	97.89	97.28	95.82	92.80	6.99	2.89	27.02	56.28
DenseNet-169	97.70	97.18	95.79	92.56	12.56	3.43	48.51	61.30
DenseNet-201	97.52	97.10	95.65	92.43	18.18	4.39	70.21	66.09
DenseNet-SC	98.29	97.64	96.80	93.26	3.88	0.72	14.94	15.22

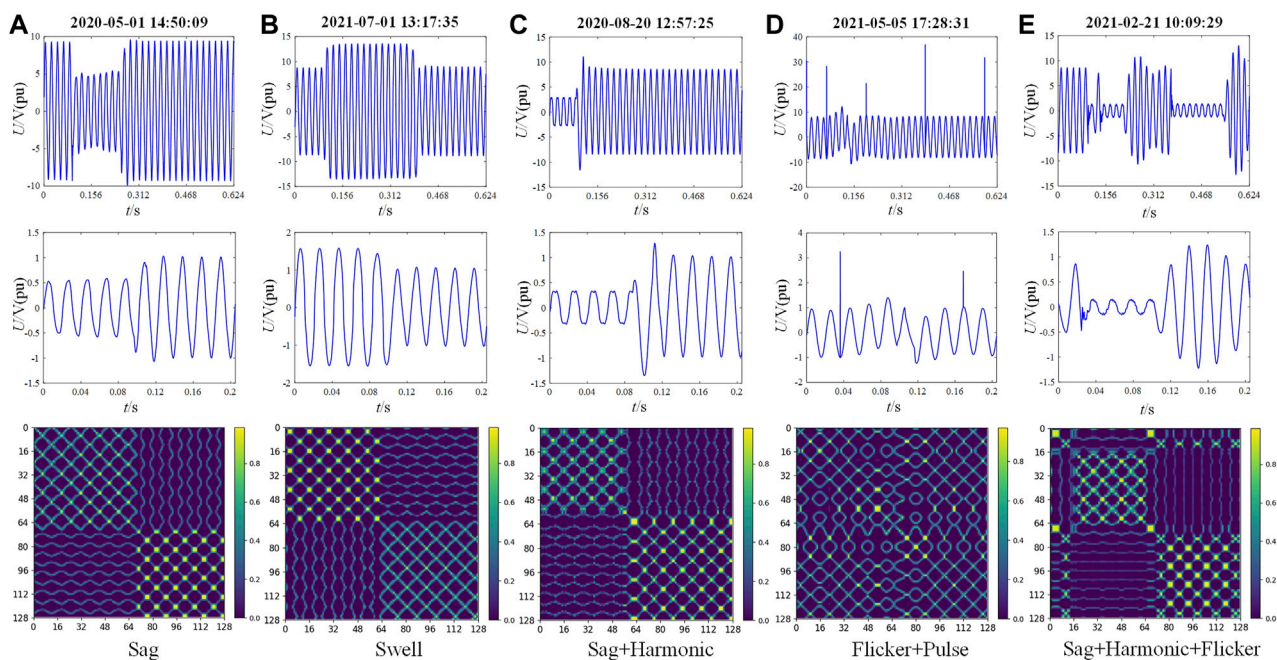


FIGURE 10
Real-world field measured waveform and MTF conversion diagram of typical PQDs event.

- (4) Classification Module: This module consists of an adaptive average-pooling layer, two fully connected layers, and a Softmax classifier. Firstly, the feature maps enhanced by CBAM are converted into one-dimensional feature vectors using adaptive average pooling, which is then input into the fully connected layer. Finally, the feature information output from the fully connected layer is input into the Softmax classifier. Softmax function is used to calculate the probability value of each PQDs type corresponding to the MTF feature image, and the category of the maximum value is output as the classification result to realize the recognition of PQDs type.

4 Simulation analysis

4.1 PQDs visualization dataset generation

According to IEEE Std 1159–2019, 34 composite disturbances are generated by combining nine mathematical models of basic disturbance signals. These composite disturbances included 18 double disturbances, 11 triple disturbances, and 5 quadruple disturbances. The random parameters of PQDs including fluctuation amplitude and duration are based on Table 1. Additionally, in alignment with Section 2.3's outlined signal sampling parameters, 1,000 samples are generated for each disturbance with uniform amplitude and phase distribution. These samples are produced at varying signal-to-noise ratios (SNRs) of 0 dB, 20 dB, 30 dB, and 50 dB to better simulate real-world PQDs scenarios. Through the visual conversion process of PQDs signal in Section 2.3, MTF is used to map samples into two-dimensional images, and a total of 43 types of PQDs visualization dataset is constructed.

This paper presents the development of a PQDs recognition model utilizing the PyTorch deep learning framework in Python 3.9. The experimental environment uses AMD Ryzen 3,970X @ 3.70 GHz CPU, 128 GB RAM, and NVIDIA RTX 3090 GPU. The cross-validation technique is used in the training process. In each epoch, the data from all categories are randomly arranged and divided into training, validation, and test sets according to the ratio of 6:2:2, and the optimal model is saved according to the recognition accuracy of validation sets.

4.2 Model parameter settings and evaluation criteria

When training the PQDs recognition model, the batch size is set to 64, the number of training epochs is set to 50, and the initial learning rate is set to 0.001. In order to obtain the optimal training model, the stochastic gradient optimizer (SGD, weight_decay = 0.0001, momentum = 0.9) is used to optimize the model, and the cross entropy loss function is used to calculate the loss value. To ensure the network learning efficiency and prevent overfitting, a dynamic adjustment strategy is used to update the learning rate. After 50% of the total epochs are completed, the learning rate is adjusted to the original 10%.

To evaluate the model's performance, this paper uses multiple evaluation metrics such as average recognition accuracy (Accuracy), floating-point operations (FLOPs), parameters (Params), and model size. Accuracy is calculated using the subsequent Eq. 6: The formula for calculating the spatial attention weights is as follows:

$$\text{Accuracy} = \frac{1}{N} \sum_{j=1}^n \frac{M_{jj}}{M_j} \times 100\% \quad (6)$$

TABLE 4 Recognition results of real-world field measured PQDs signal.

Serial number	Disturbance types	Recognition results of the proposed algorithm
1	Sag: 51 Groups	Sag: 45 Groups/Interruption: 2 Groups Sag + Oscillation transient: 3 Groups/Sag + Notch: 1 Group
2	Swell: 22 Groups	Swell: 22 Groups
3	Interruption: 3 Groups	Interruption: 3 Groups
4	Harmonic: 24 Groups	Harmonic: 23 Groups/Harmonic + Oscillation transient: 1 Group
5	Pulse: 5 Groups	Pulse: 5 Groups
6	Flicker: 34 Groups	Flicker: 30 Groups/Flicker + Harmonic: 3 Group/Flicker + Pulse: 1 Group
7	Sag + Harmonic: 15 Groups	Sag + Harmonic: 13 Groups/Sag: 2 Groups
8	Sag + Oscillation transient: 10 Groups	Sag + Oscillation transient: 8 Groups/Sag + Harmonic: 1 Group/Sag + Oscillation transient + Harmonic: 1 Group
9	Flicker + Pulse: 4 Groups	Flicker + Pulse: 4 Groups
10	Swell + Spike: 5 Groups	Swell + Spike: 5 Groups
11	Sag + Harmonic + Flicker: 5 Groups	Sag + Harmonic + Flicker: 5 Groups
12	Interruption + Harmonic + Oscillation transient: 3 Groups	Interruption + Harmonic + Oscillation transient: 2 Groups/Sag + Harmonic: 1 Group

where N is the total number of categories, M_j is the number of the j -th category, and M_{jj} is the number of correct recognition for the j -th category.

Params and FLOPs in CNNs are widely used to evaluate the complexity of the model (Paoletti et al., 2021). The formulas is calculated using the subsequent equation: for calculating Params and FLOPs in the convolution layer and the fully connected layer are as Eqs 7–10:

$$P_{CNN} \sim O\left(\sum_{l=1}^d (n_{l-1} \cdot s_l^2 + 1) \cdot n_l\right) \quad (7)$$

$$C_{CNN} \sim O\left(\sum_{l=1}^d n_{l-1} \cdot s_l^2 \cdot n_l \cdot m_l^2\right) \quad (8)$$

$$P_{FCN} \sim O\left(\sum_{l=1}^d (n_{l-1} + 1) \cdot n_l\right) \quad (9)$$

$$C_{FCN} \sim O\left(\sum_{l=1}^d n_{l-1} \cdot n_l\right) \quad (10)$$

where l is the index of a convolutional layer, and d is the depth (number of convolutional layers). n_l is the number of filters (also known as “width”) in the l -th layer. n_{l-1} is also known as the number of input channels of the l -th layer. s_l is the spatial size (length) of the filter. m_l is the spatial size of the output feature map.

4.3 Experimental results and analysis

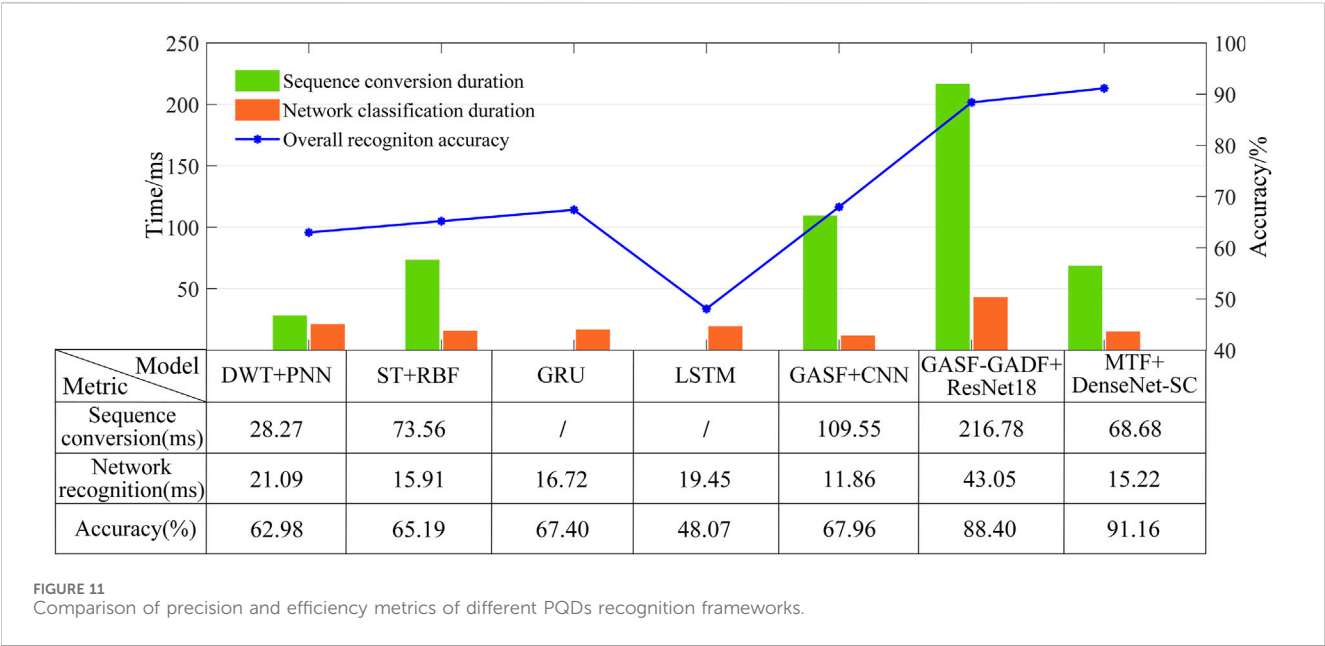
In order to verify the performance of the improved DenseNet-SC model designed in this paper in recognizing PQDs types, the improved model is compared with the DenseNet-S model and six mainstream deep learning classification networks under the same

experimental environment and SNR conditions, so as to verify the effectiveness and superiority of the improved model.

4.3.1 DenseNet-SC model capability assessment

The improved DenseNet-SC model and the DenseNet-S model without CBAM module are trained with the same dataset and experimental environment. After 50 epochs of training, the recognition accuracy and loss value changes of the verification set in the training process of the model are obtained, as shown in Figure 6. At the same time, comparison results of various models' recognition accuracy on the test set can be obtained under different SNR environments, among which the overall recognition accuracy of DenseNet-S is 95.72%, 95.06%, 93.92%, and 89.72% under no noise, 50 dB, 30 dB, and 20 dB environments, respectively. In the identical SNR environments, the overall recognition accuracy of DenseNet-SC is 98.29%, 97.64%, 96.80%, and 93.26%, respectively.

As can be seen from Figure 6, the accuracy and convergence rate of the improved DenseNet-SC model are significantly better than that of the DenseNet-S model during the epochs. Once it reaches the state of convergence, the accuracy of the former is stable at about 98%, and the loss value is stable at about 0.07, while the accuracy and loss value of the latter are oscillating at about 95% and 0.17, respectively. This suggests that the addition of CBAM module to the DenseNet-S model effectively improves the extraction of detailed features in the MTF images that correspond to PQDs signals, so as to grasp the feature information in the input image more accurately. By comparing the overall recognition results of the two models on the test set, it can be seen that the DenseNet-SC model also performs better than DenseNet-S in PQDs recognition under four noise environments: no noise, 50 dB, 30 dB, and 20 dB. The recognition accuracy is improved by 2.57, 2.58, 2.88, and 3.54 percentage points respectively, which further indicates that adding CBAM attention mechanism between the dense connection learning module and the classification module of model is more helpful for the



model to focus on key feature information, so as to improve the recognition performance.

The Grad-Cam method (Selvaraju et al., 2020) showcases the degree to which different network modules concentrate on image features via a heatmap. The heatmap's colorbar value indicates the degree of network focus, with higher values indicating a greater level of focus. The class activation heatmaps of the output features of some samples before and after CBAM module by the improved model are shown in Figure 7. The figure indicates that, regardless of basic, double, triple, or quadruple disturbances, the image areas of focus are more precise and comprehensive in capturing key texture features in the MTF images when passing through the CBAM module than when passing through the dense connection learning module. The reason for this is that the deeper features of feature extraction module are more related to global information, and the CBAM module assigns more weight coefficients to the network from a global viewpoint. Therefore, deep global features can be better extracted and learned through CBAM module, which enables the focused feature areas to cover the MTF texture pattern areas with key feature information more comprehensively.

4.3.2 Model noise resistance performance evaluation

In noisy environments, PQDs signals may become distorted, potentially interfering with the model's judgment of their real class and causing a reduction in the network's recognition precision. To evaluate the anti-noise ability of the improved model, combined with the model recognition performance under different SNR environments in the previous section, the recognition effect on various types of disturbances is tested respectively in the 20 dB SNR environment. The corresponding confusion matrix of the test results is shown in Figure 8, where the rows indicate the actual disturbance labels while the columns reflect the model's recognition outcomes.

The test results show that under the 20 dB strong noise environment, the recognition accuracy of some complex disturbances containing oscillatory transient, harmonic, flicker

and spike, is below average. Among them, some disturbances are easily confused with complex disturbances containing notch and spike (C1 → C5/C6, C3 → C5/C6, C14 → C17). Some double disturbances with harmonic can be easily identified as triple disturbances (C2 → C8, C24 → C28, C32 → C36) and the quadruple disturbances with oscillatory transient are easily confused with the triple disturbances without oscillatory transient (C9 → C10, C20 → C21), resulting in a significant decline in recognition accuracy. On the other hand, the overall recognition accuracy of the improved model can still reach 93.26% even under 20 dB strong noise environment, which shows good recognition performance and noise robustness.

4.3.3 Comparison of recognition precision performance among different deep learning models

To further verify the precision performance of DenseNet-SC model in recognizing PQDs, we performed comparative experiments with six mainstream deep learning classification networks: GoogleNet (Inception V3), Vgg-16, ResNet-18, and DenseNet-121/169/201 under equivalent conditions of MTF dataset and experimental settings. These classification networks utilize the corresponding framework structure and parameters setting from the original papers. Figure 9 shows the comparison curves of recognition accuracy, while Table 3 presents the recognition accuracy results on the test set. It is important to note that MTF images need to be upsampled to 299 × 299 and 224 × 224 respectively before being input into the networks so as to adapt to the pre-trained GoogleNet and the other five classification networks, while the proposed DenseNet-SC does not require any prior information.

As shown in Figure 9A, compared with the three DenseNet series classification networks, although the initial recognition accuracy of the improved model is relatively low, its accuracy continues to increase and the rising trend is very stable with the increase of epochs, and there is no

obvious fluctuation. In the 16th epoch, the initial synchronization with other models is achieved, and the surpassing is completed in the later stage. In addition, as can be seen from Table 3, the recognition accuracy of the improved model under different SNR environments is superior to the other three models in terms of accuracy performance verification of the test set. In summary, although DenseNet-SC model is lightweight in the network architecture, the increase in growth rate and the injection of CBAM attention module strengthen the propagation of feature information and the extraction of key features, improve the information flow and feature capture capability of the whole network, and optimize its recognition precision performance for PQDs.

According to Figure 9B; Table 3, GoogleNet underperforms on the MTF dataset. Its recognition accuracy is lower than that of other models in both training and testing phases. Compared with VGG-16 and ResNet-18 models, on the one hand, although the improved model exhibits subpar performance during the early training phase, the recognition accuracy gradually exceeds that of two models with the increase of epochs, and finally stabilizes around 98%. On the other hand, in terms of test performance, the overall recognition accuracy of the improved model under the SNR of 0, 20, 30, and 50 dB is improved by 2.01/0.44, 1.97/0.41, 2.43/0.37, and 2.47/0.35 percentage points compared with the two models, respectively. These results demonstrate the improved model's superiority in terms of PQDs recognition precision and noise resistance.

4.3.4 Comparison of recognition efficiency performance among different deep learning models

Considering the high requirement of PQDs recognition efficiency for power system health status monitoring, the size of recognition model and its operational efficiency are important evaluation metrics to judge its performance. The size and operational performance of different recognition models are shown in Table 3, where time refers to the average time of taking the model to recognize a sample image from the test set over 100 tests.

As can be seen from Table 3, when compared with ResNet-18 and the three DenseNet series models, the recognition accuracy of Densenet-SC model improves relatively little, but the number of Params, FLOPs, model size and recognition time are greatly reduced. Among them, the Params and the model size of DenseNet-SC are only 55.51% of DenseNet121 model, which is the lightest model among the six comparison models, while the FLOPs and the recognition time to recognize an image are less than 40% of ResNet-18 model, which has the highest operational efficiency among the comparison models. This can not only effectively reduce the memory ratio and computing power, but also help to improve the training and testing efficiency of the model, so as to realize the PQDs recognition operation performance upgrade of the model. On the other hand, compared with the GoogleNet and VGG-16 models, the proposed model has achieved significant optimization in terms of recognition accuracy, model size and computing performance. In addition, due to the influence of input image size and its own network structure, GoogleNet's efficiency of recognizing an image is lower than that of VGG-16. To sum up, the DenseNet-SC model designed in this paper not only improves the network recognition precision, but also realizes the lightweight and high efficiency of the network, making it able

to be deployed on the hardware terminal equipment with small storage capacity and low computing performance configuration, which provides the possibility for further exploring the construction of mobile PQDs recognition system.

5 Real-world field measured signals analysis

5.1 Practical effectiveness analysis of the proposed PQD recognition method

To assess the effectiveness of the proposed method for practical engineering use, real-world field measured PQDs signal data is utilized to test its recognition performance. The data used for testing are gathered from a power quality monitoring device in a 10 kV substation located in the southern region of Jiangsu Province, China. The data collection period ranges from March 2020 to August 2021, comprising a total of 181 sets of samples. The recording device has a sampling frequency of 12.8 kHz and the signal lasted for 0.624 s. To conform with IEEE Std1159-2019 and the input requirements of the proposed recognition network, the 10-cycle typical data of each set of signals is intercepted and normalized as the input of signal visualization module. The real-world field measured waveforms of PQDs typical events and their MTF conversion images are shown in Figure 10. It can be seen that the MTF image corresponding to each disturbance measured signal still have clear and easily distinguishable texture pattern features, which are basically consistent with the theoretical MTF image features. Therefore, it can be seen that the MTF conversion mode also has good feature expression ability for the real-world field measured signal data. MTF samples corresponding to all measured signals are identified by the DenseNet-SC optimal training model, and the results are shown in Table 4.

As can be seen from Table 4, in terms of sag recognition, 2 groups are recognized as interruption due to the too low sag amplitude of signal sequence, and 4 groups are recognized as sag + oscillatory transient/notch due to environmental noise interference. In terms of flicker recognition, 3 groups are detected to contain harmonic components due to the influence of noise, and 1 group is identified to contain pulse disturbance due to the presence of serious noise point. In terms of double disturbance recognition, the two groups of sag + harmonic are identified as sag. Further analysis shows that the harmonic content in the measured signal is small and its total proportion is relatively small, resulting in label loss. In addition, a group of triple disturbance signals (interruption + harmonic + oscillatory transient) are recognized as sag + harmonic because the signal sag amplitude is near the critical value and the oscillatory amplitude is relatively small. In summary, due to the mismatch between the real-world field measured signal data and the simulation signal data simulated by mathematical model, the recognition accuracy of some measured disturbance types decreases compared with the simulation results, but on the whole, the measured disturbance signal types can be effectively recognized with a high precision of 91.16%. Therefore, the effectiveness and reliability of the proposed method in engineering practice are verified.

5.2 Comparative performance analysis with existing PQDs recognition frameworks

To further verify the comprehensive performance advantages of the proposed method in terms of efficiency and accuracy, six existing PQDs identification algorithms that have achieved excellent performance are selected as benchmark algorithms for comparative analysis with the proposed algorithm on the real-world field measured dataset. The benchmark algorithms are divided into machine learning algorithms and deep learning algorithms.

The machine learning algorithms include DWT + PNN (Khokhar et al., 2017) and ST + RBF (Wang et al., 2018), while the deep learning algorithms include the one-dimensional deep learning algorithms GRU (Deng et al., 2019) and LSTM (Xu et al., 2022), which directly take one-dimensional signals as model inputs, and the two-dimensional deep learning algorithms GASF + CNN (Zheng et al., 2021) and GASF-GADF + ResNet18, which convert one-dimensional signals into two-dimensional visualization images and then combine advanced image recognition algorithms for PQDs classification.

To ensure the fairness of various algorithms in the PQDs recognition performance test, we process the data and train the classifier models based on the same original one-dimensional signal dataset under the same experimental environment. Among them, the original dataset is divided according to the same dataset partitioning ratio as in this paper, the classifier model structure is consistent with that in the original paper, and the hyperparameter settings are consistent with that in this paper. After the training of various algorithm models and saving the optimal models, they are applied to the measured signal dataset and the comprehensive performance is compared with the proposed method. The results are shown in Figure 11. Among them, the visual conversion and network classification time are the average time obtained for a signal sample after 100 consecutive tests. Sequence conversion time and network classification time refers to the average time used to process and recognize a signal sample in the dataset, respectively.

According to the test results, the overall accuracy performance of machine learning algorithms is poor. This is because this algorithm type relies heavily on expert knowledge when manually extracting features, and its generalization performance is greatly limited when facing measured signals with more complex changing states. In addition, machine learning models often have shallow depth, which leads to their lack of deep feature capture ability, thereby causing a decline in PQDs recognition accuracy performance. One-dimensional deep learning algorithms directly take one-dimensional sequence data as model input without any data processing, so they are superior to other algorithms in efficiency performance. However, their accuracy performance is limited due to its limitation in information mining of complex time series data. With the deep feature mining ability and strong generalization of advanced image recognition algorithms, the accuracy performance of two-dimensional deep learning algorithms is better than that of traditional machine learning algorithms and one-dimensional deep learning algorithms. Among them, due to the limited temporal and spatial features of the PQDs signals extracted by single-channel GASF in the coding process, and the shallow depth of CNN network, the deep-level semantic information of the image cannot be captured, thus the overall recognition accuracy of GASF + CNN

is only 67.96%. Compared with single-channel GASF, double-channel GASF-GADF can provide more abundant PQDs time-series feature information. However, real-world disturbance signal aliases environmental noise and its state changes are more complex, which makes the GAF visualization process requiring coordinate conversion and matrix coding inevitably lose some important feature information. It interferes with the classification network to capture important feature information, and then reduces the recognition performance of the model, so its accuracy is also limited. Compared with GAF, MTF is more resistant to interference and noise through the introduction of quantile division, and the encoded image obtained by conversion can retain more comprehensive temporal and spatial feature information. In terms of recognition efficiency, the visual conversion efficiency of the proposed method is much higher than other two methods, because the MTF conversion process is very simple and computationally small, and it does not need to carry out complex coordinate conversion and matrix coding. In addition, due to the lightweight structural design, the model recognition efficiency is higher than that of the larger ResNet-18, but it is lower than that of CNN, because CNN has a simpler structure (the number of parameters and the model size are 2.18 M and 8.32 MB, respectively). However, CNN has obvious defects in the recognition accuracy performance. In summary, the experimental results of measured signals prove that the MTF + DenseNet-SC method proposed in this paper has better comprehensive performance in PQDs recognition than other algorithms.

6 Conclusion

To meet the recognition efficiency and precision requirements of massive and complex PQDs events under the background of new power system, and give full play to the deep sensing capability of DenseNet and the simple conversion and anti-interference capability of MTF, a novel PQDs recognition method based on MTF and improved DenseNet is proposed in this paper. The specific advantages are as follows:

- 1) MTF can effectively extract the disturbance characteristics of one-dimensional signals, and has certain anti-interference ability. The two-dimensional images after visualization can effectively express the sample disturbance information.
- 2) A DenseNet-S lightweight network structure is designed, and by introducing an attention mechanism, the improved model can capture the texture pattern features of each MTF sample more accurately and comprehensively. The recognition accuracy of 43 PQDs types can reach 98.29% and 93.26% respectively under noise-free and 20 dB strong noise environment, which effectively improves the PQDs recognition precision and noise robustness.
- 3) The improved DenseNet-SC model has obvious advantages over six mainstream deep learning image classification models, such as ResNet-18 and DenseNet-121, in terms of model size, Params, FLOPs and recognition time. While improving recognition precision, the lightweight of model and the high efficiency of data analysis are realized. In addition, this paper constructs 12 types of real-world field measured PQDs signal dataset to comprehensively test the recognition performance of the

proposed method. The test results verify the effectiveness and superiority of the proposed method in the accuracy and efficiency of real-world field measured PQDs recognition, which can meet the classification accuracy and efficiency requirements of massive and complex PQDs events in engineering applications.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LZ: Conceptualization, Formal Analysis, Investigation, Methodology, Validation, Writing—original draft. SG: Data curation, Investigation, Resources, Validation, Writing—review and editing. YL: Resources, Supervision, Validation, Writing—review and editing. CZ: Resources, Software, Supervision, Writing—review and editing.

References

- Ahmadi, A., and Tani, J. (2019). A novel predictive-coding-inspired variational RNN model for online prediction and recognition. *Neural Comput.* 31 (11), 2025–2074. doi:10.1162/neco_a_01228
- Cui, C. H., Duan, Y. J. V., Hu, H. I., Wang, L., and Liu, Q. (2022). Detection and classification of multiple power quality disturbances using stockwell transform and deep learning. *IEEE Trans. Instrum. Meas.* 71, 1–12. doi:10.1109/TIM.2022.3214284
- Deng, Y., Wang, L., Jia, H., Tong, X., and Li, F. (2019). A sequence-to-sequence deep learning architecture based on bidirectional GRU for type recognition and time location of combined power quality disturbance. *IEEE Trans. Industrial Inf.* 15 (8), 4481–4493. doi:10.1109/TII.2019.2895054
- He, C. J., Li, K. C., and Yang, W. W. (2023). Power quality compound disturbance identification based on dual channel GAF and depth residual network. *Power Syst. Technol.* 47 (1), 369–376. doi:10.13335/j.1000-3673.pst.2022.0644
- Huang, G., Liu, Z., Van, D. M., and Weinberger, K. Q. (2017b). Densely connected convolutional networks. *Proc. CVPR*, 4700–4708. doi:10.1109/CVPR.2017.243
- Huang, J. M., Ju, H. Z., and Li, X. M. (2016). Classification for hybrid power quality disturbance based on STFT and its spectral kurtosis. *Power Syst. Technol.* 40 (10), 3184–3191. doi:10.13335/j.1000-3673.pst.2016.10.036
- Huang, N. T., Peng, H., and Cai, G. W. (2017a). Feature selection and optimal decision tree construction of complex power quality disturbances. *Proc. CSEE* 37 (3), 776–785. doi:10.13334/j.0258-8013.psee.160108
- Jyoti, S., Basanta, K. P., and Prakash, K. R. (2021). Power quality disturbances classification based on Gramian angular summation field method and convolutional neural networks. *Int. Trans. Electr. Energy Syst.* 31 (12), e13222. doi:10.1002/2050-7038.13222
- Khokhar, S., Zin, A. A., Memon, A. P., and Mokhtar, A. S. (2017). A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network. *Measurement* 95 (2017), 246–259. doi:10.1016/j.measurement.2016.10.013
- Li, D. Q., Mei, F., Zhang, C. Y., Sha, H. Y., Zheng, J. Y., and Li, T. R. (2020). Deep belief network based method for feature extraction and source identification of voltage sag. *Automation Electr. Power Syst.* 44 (04), 150–160. doi:10.7500/AEPS20190306004
- Li, R. J., W. Y., Wu, Q., Nilanjan, D., Rubén, G. C., and Shi, F. Q. (2022). Emotion stimuli-based surface electromyography signal classification employing Markov transition field and deep neural networks. *Measurement* 189 (2022), 110470. doi:10.1016/j.measurement.2021.110470
- Paoletti, E. M., Haut, M. J., Tao, X. W., Plaza, J., and Plaza, A. (2021). FLOP-reduction through memory allocations within CNN for hyperspectral image classification. *IEEE Trans. Geoscience Remote Sens.* 59 (7), 5938–5952. doi:10.1109/TGRS.2020.3024730
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-CAM: visual explanations from deep networks via gradient-based localization. *Proc. ICCV*, 618–626.
- Sindi, H., Nour, M., Rawa, M., Öztürk, Ş., and Polat, K. (2021). An adaptive deep learning framework to classify unknown composite power quality event using known single power quality events. *Expert Syst. Appl.* 178, 115023. doi:10.1016/j.eswa.2021.115023
- Std (2019). IEEE recommended practice for monitoring electric power quality. *IEEE* 29, 240. 01-Power transmission and distribution networks in general.
- Tang, Q., Qiu, W., and Zhou, Y. C. (2020). Classification of complex power quality disturbances using optimized S-transform and kernel SVM. *IEEE Trans. Industrial Electron.* 67 (11), 9715–9723. doi:10.1109/TIE.2019.2952823
- Wang, F., Quan, X. Q., and Ren, L. T. (2021). Review of power quality disturbance detection and identification methods. *Proc. CSEE* 41 (12), 4104–4120. doi:10.13334/j.0258-8013.psee.201261
- Wang, H., Wang, P., Liu, T., and Zhang, B. (2018). Power quality disturbance classification based on growing and pruning optimal RBF neural network. *Power Syst. Technol.* 42 (8), 2408–2415. doi:10.13335/j.1000-3673.pst.2017.0663
- Wang, S., and Chen, H. (2019). A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. *Appl. energy* 235, 1126–1140. doi:10.1016/j.apenergy.2018.09.160
- Wang, Z., and Oates, T. (2015). Spatially encoding temporal correlations to classify temporal data using convolutional neural networks. *J. Comput. Syst. Sci.*, 07481. *arXiv*: 1509. doi:10.48550/arXiv.1509.07481
- Woo, S., Park, J. P., Lee, J. Y., and Kweon, I. S. (2018). CBAM: convolutional block attention module. *Proc. ECCV*, 3–19. doi:10.1007/978-3-030-01234-2_31
- Wu, J. Z., Mei, F., Zhen, J. Y., Zhang, C. Y., and Miao, H. Y. (2022). Recognition of multiple power quality disturbances based on modified empirical wavelet transform and XGBoost. *Trans. China Electrotech. Soc.* 37 (1), 232–243. doi:10.19595/j.cnki.1000-6753.tces.201363
- Xu, W., Duan, C., Wang, X., and Dai, J. (2022). Power quality disturbance identification method based on improved fully convolutional network. *Proc. IEEE Asia Conf. Energy Electr. Eng.*, 1–6. doi:10.1109/ACEEE56193.2022.9851835
- Yan, J. L., Kan, J. M., and Luo, H. F. (2022). Rolling bearing fault diagnosis based on Markov transition field and residual Network. *Sensors* 22 (10), 3936. doi:10.3390/s22103936
- Yin, B. Q., Chen, Q. B., and Li, B. (2021). A new method for identification and classification of power quality disturbance based on modified Kaiser window fast S-transform and LightGBM. *Proc. CSEE* 41 (24), 8372–8383. doi:10.13334/j.0258-8013.psee.210743
- Zheng, W., Lin, R., and Wang, J. (2021). Power quality disturbance classification based on GAF and a convolutional neural network. *Power Syst. Prot. Control.* 49(11), 97–104. doi:10.19783/j.cnki.pspc.200997

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the State Grid Jiangsu Electric Power Co., Science and Technology Project (J2022093).

Conflict of interest

Authors LZ, SG, YL, and CZ were employed by State Grid Suzhou Power Supply Company.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



OPEN ACCESS

EDITED BY

Bo Wang,
Wuhan University, China

REVIEWED BY

Ge Cao,
Xi'an University of Technology, China
Guangchen Liu,
Inner Mongolia University of Technology, China

*CORRESPONDENCE

Jie Hu,
✉ Jhon.Hu@outlook.com

RECEIVED 16 November 2023

ACCEPTED 28 December 2023

PUBLISHED 12 January 2024

CITATION

Hu J, Xu G, Qi L and Qie X (2024), Knowledge reasoning in power grid infrastructure projects based on deep multi-view graph convolutional network.
Front. Energy Res. 11:1339416.
doi: 10.3389/fenrg.2023.1339416

COPYRIGHT

© 2024 Hu, Xu, Qi and Qie. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Knowledge reasoning in power grid infrastructure projects based on deep multi-view graph convolutional network

Jie Hu^{1*}, Gang Xu¹, Lizhong Qi² and Xin Qie¹

¹School of Electrical and Electronic Engineering, North China Electric Power University, Beijing, China,

²State Grid Economic and Technological Research Institute Co Ltd., Beijing, China

With the rapid development of power grid infrastructure, especially the increasing number of ultra-high voltage (UHV) projects, knowledge extracted from historical engineering data is collected and can be potentially used to assist in the review of power transmission and transformation projects. However, conventional knowledge modeling and knowledge reasoning methods cannot meet the current needs of power grid construction. In this paper, considering the more supernumerary and distinctive information brought by multi-view data which could be beneficial for feature representation and knowledge reasoning from the constructed knowledge base, a multi-view graph convolutional network (GCN) based on knowledge graph is proposed to make classification for power grid infrastructure projects. Specifically, several views are constructed based on attribute information of a knowledge graph. In addition, a Haar convolution-based pooling mechanism is employed to capture the structural features represented by a chain of subgraphs. And then an aggregator that combines both attribute and structural information is used to classify UHV projects. Results from both UHV and NCI-1 datasets indicate that our proposed method is more has higher accuracy and generalization ability.

KEYWORDS

review of power transmission and transformation, knowledge graph, graph classification, graph convolutional network, knowledge reasoning

1 Introduction

With the proposal of carbon peaking and carbon neutralization (J. Liu et al., 2021; Luo et al., 2023; Ren et al., 2021), a new power system based on clean energy consumption has become an important part of achieving dual-carbon goals. In this context, photovoltaic and wind power from the western region of China have become important power sources with a total scale of 450 million kilowatts and increasing. Due to the long distance of electricity transmission, UHV projects are expected to enter a larger-scale construction stage. Unlike conventional power transmission and transformation projects, UHV projects are larger in construction scale with multiple companies involved, which lead to a huge amount of accumulated historical data in terms of quantity and complexity. With the development of digital technology and artificial intelligence and the deep integration of technology and information technology in the field of power grid engineering construction, the concept of digital infrastructure has emerged. As an important part of engineering construction management and control, engineering review is still in an inefficient mode that relies mainly on expert experience and offline review, which makes it difficult to meet the analysis

requirements of the current ever-expanding construction scale of power transmission and transformation projects. Therefore, it is urgent to build an auxiliary review system with knowledge storage and reasoning components based on historical data of power transmission and transformation projects, which leads to an improvement of the digitalization level and the efficiency of power grid infrastructure project review.

At present, the studies of auxiliary review platforms of power transmission and transformation projects are still in the early stages. Browser/Server(B/S) architecture and SQL server databases are two key technologies of these systems that have already been put into the production phase. Built for various voltage levels and different workflows, three key functions are implemented such as data entry, key factor extraction, and data searching (Huang, 2018; Li et al., 2021). Although the application of these platforms has simplified the review process and improved the efficiency of data processing, there are three difficulties yet not be solved: 1) the knowledge extracted from heterogeneous data is stored in relational databases and their connection relationships have not been modeled; 2) knowledge reasoning is difficult based on existing models and data storage methods; 3) knowledge extraction methods are heavily based on expert experiences and human labor. Therefore, knowledge graph technology, which has the advantages of high scalability, high query efficiency, and good visualization, has become one of the excellent technologies to choose from in building the next-generation power transmission and transformation engineering auxiliary review platform.

The knowledge graph is a structured semantic knowledge base that integrates knowledge extraction, data storage, reasoning, and analysis capabilities (Ji et al., 2022). It has been widely used in many fields (Yang et al., 2022; Zou and Lu, 2022; Wu et al., 2023). At present, the application of knowledge graphs in the power field is mainly oriented to aspects such as dispatching, operation and maintenance, and fault handling, and has achieved good results (Pu et al., 2021; Tian et al., 2022; Liu et al., 2023). These researches mainly focus on construction methods such as named entity recognition and relation extraction algorithms of different domains. However, the methodologies and applications of knowledge reasoning technology based on domain knowledge graphs in electric power systems are still in an early stage. Knowledge graphs can be represented as semantic triples or attributed networks (Gao et al., 2023), which are all non-European structural graphs. Traditional deep learning methods such as convolutional neural network(CNN) and recurrent neural network(RNN) cannot be used in such scenarios because representations of graph-structured data is generally irregular, therefore a new deep learning mechanism is needed to process graph structure. In this context, the graph neural network (GNN) is proposed to get the latent representations embedded from nodes, attributes, and structural information of graphs. Compared to the basic network structure of the neural network, the fully connected layer (MLP), which multiplies the feature matrix by the weight matrix, the graph neural network considers structural information, and adds an adjacency matrix as input. With the development of the study, several GNN variants have been proposed. For the first time, graph convolution networks(GCN) introduce convolution operations in image processing to process graph-structured data (Kipf and Welling, 2016). Various experiments show the effectiveness of

GCN due to the ability to encode the structural information of the graph. However, the shortcomings of GCN are also obvious: 1) GCN needs to put the entire graph data into memory and graphics processing unit(GPU), which requires high-performance equipment when dealing with large graphs. 2) GCN has high computational complexity due to the eigendecomposition operation of graph Laplacian. To handle the problems mentioned above, graph sample and aggregate(GraphSAGE) are proposed (Hamilton et al., 2017). GraphSAGE is an inductive learning framework. In practice, it only retains the training sample to the edge of the training sample during training, and then includes the two major steps which are sample and aggregate. Then, to solve the problem that GNN does not take into account the different importance of different neighbor nodes when aggregating neighbor nodes, graph attention networks (GAT) take the idea of Transformer and introduce the masked self-attention mechanism (Veličković et al., 2017). When calculating the representation of each node in the graph, different weights are assigned to neighbor nodes based on their different characteristics.

However, there are different types of attributes of a node in the knowledge graph. For instance, a transformer could have three attributes such as device model, quantity, and rated voltage. Both of the latter two are usually numeric properties and the first one is a combination of letters and numbers. Conventional GNN-based knowledge modeling and reasoning methods take these attributes as a single matrix, which may lead a confusion about attribute characteristics (Peng et al., 2020). To solve the knowledge reasoning problem in power grid infrastructure projects, a deep multi-view graph convolutional network is introduced considering both the attribute and structural information. This model mainly contains 3 components: a multi-view graph encoder, an aggregator, and classification module. The key contributions are as follows:

- 1) Multi-view information is used separately in knowledge reasoning of UHV projects. It can provide a more accurate feature embedding than a single-view latent representation.
- 2) A more effective pooling mechanism based on the Haar convolution method is introduced considering the structural information of the knowledge graph in the UHV projects.
- 3) Based on the construction of knowledge graphs in the UHV projects domain and the proposition of a deep multi-view graph convolutional network, the problem of knowledge reasoning on project classification is solved, which lays the foundation for the downstream application of the power transmission and transformation auxiliary review platform.

2 Problem formulation

The knowledge graph of UHV projects can be divided into two categories based on the attributes of the nodes and the structural information of the graph. Specifically, nodes of knowledge graph can be seen as nodes in the attribute matrix, and the structure information can be modeled by adjacency matrix. As shown in Figure 1, three views are generated based on different attribute types, with each view representing a single aspect of the node attributes, such as rated voltage, device model, and quantity. Several subgraphs are also produced based on the structural information of the graph,

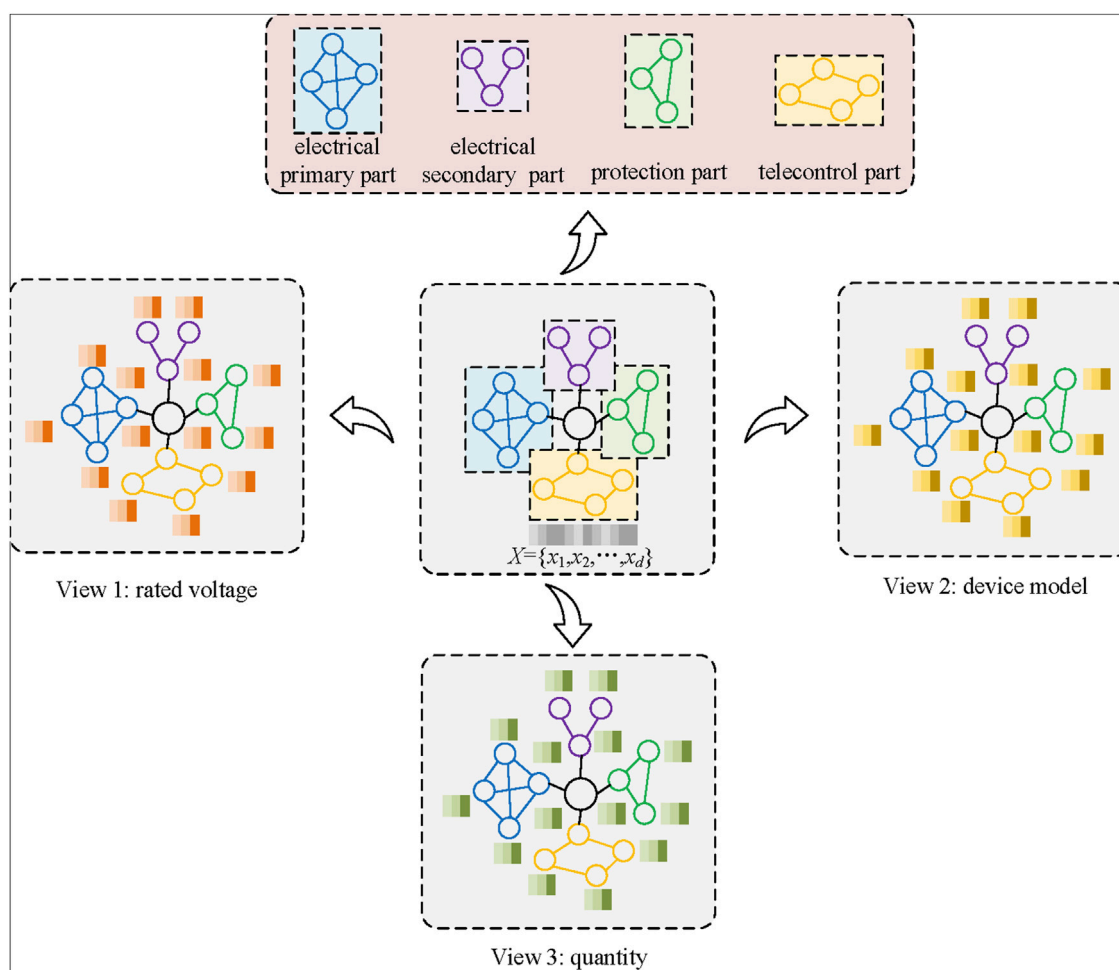


FIGURE 1
A UHV example to illustrate the multiple views and structural divide.

such as the electrical primary part, electrical secondary part, protection part, telecontrol part, and so on.

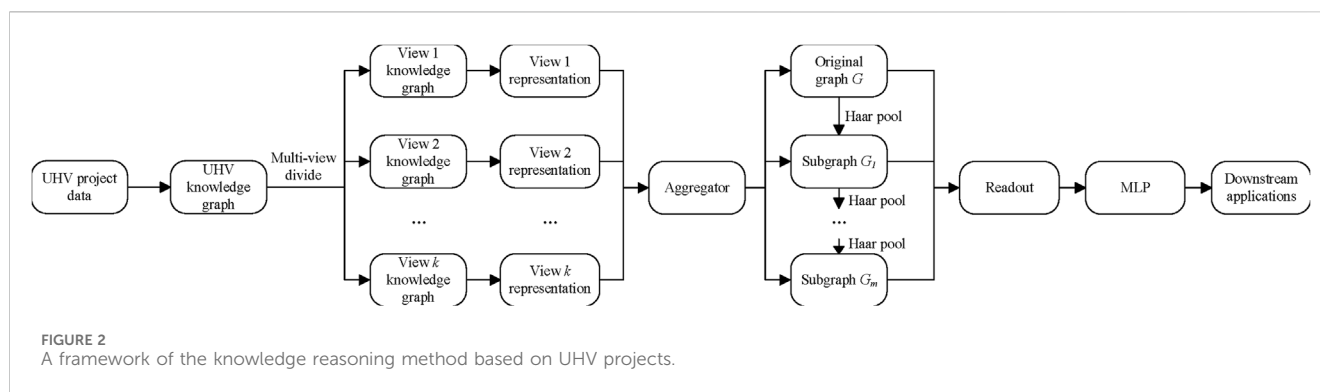
Following the commonly used notations, a knowledge graph is denoted as $G(V, A, X)$, where V denotes the set of nodes, A denotes the adjacency matrix and X denotes the attribute matrix. Thus, a multi-view graph of the graph G can be represented as $G^{mg}=(V, A, X^{mg})$, where $X^{mg} \in \mathbb{R}^{n \times d^{mg}}$ is a submatrix of X and $n = |V|$ is the number of nodes, d^{mg} contains k distinct attribute features of the graph. As is illustrated above, a multi-view graph has the same structure as the original graph which V and A remain unchanged, while the attributes of the nodes are part of X . On the other hand, a subgraph of G can be denoted as $G^s=(V^s, A^s, X)$, where V^s and A^s are the submatrix of the original V and A , representing a divide from the structural aspect.

Our purpose is to classify the knowledge graph subgraphs, which is a fundamental function for the review of power transmission and transformation projects. Therefore, a function to generate a probability of each graph should be learned with both G^{mg} and G^s as input, which could be represented by the following Equation (1), where Z denotes the probability of the graph, f is a trainable function of the deep multi-view graph convolutional network, X_{mg} and X_s denotes the embeddings of G^{mg} and G^s respectively.

$$Z = f(X_{mg}(G^{mg}), X_s(G^s)) \quad (1)$$

3 Framework and methodologies

Considering both the attribute and structural information of the UHV knowledge graph, the knowledge reasoning framework mainly takes five steps. Firstly, with the analysis of historical structured, semi-structured, and unstructured data, a UHV knowledge graph is constructed. Secondly, considering the attribute information of the knowledge graph, a series of multi-view graphs are generated and GCN is leveraged to learn the latent representation of each graph. Thirdly, an aggregator is used to combine multi-view representations. Fourthly, a hierarchical Haar graph pooling method is adopted to replace the graph Laplacian-based GNN considering the structural information of the graph. Finally, a unified representation of the knowledge graph is generated and ready for graph classification and other downstream applications. The framework of the knowledge reasoning method is shown in Figure 2. More details will be elaborated in the following sections.



3.1 Knowledge graph construction based on UHV projects

Data from the UHV projects mainly contains three parts: structural, semi-structural, and non-structural data. Project investment table, equipment list are typical structural data. Equipment inventory is semi-structural data because the properties can be put into the same cell and break the structural feature of the data. Preliminary design instructions, feasibility study reports are some of the non-structural data, which account for a large proportion. The knowledge graph is a structured semantic knowledge base used to describe concepts in the physical world and their relationships in symbolic form. In this section, a two-stage method is proposed to construct a UHV knowledge graph.

1) Knowledge graph ontology construction

Ontology is a model and a pattern constraint on the data that constitutes the knowledge graph. Building an ontology in a specific domain requires cooperation with multiple experts in vertical fields. The inputs for constructing ontology include domain knowledge, terminology dictionary, experience of experts, etc. The output includes the entity categories that constitute the knowledge graph, as well as the relationships between categories, the set of attributes that entities of a specific category, and so on. There are generally three methods for constructing knowledge graph ontology: top-down, bottom-up, and a combination of the two. Among them, the top-down approach is to define the most common top-level concepts in the field and then expand downwards in sequence. It requires a thorough understanding of a specific domain both in the business aspect and the data aspect as well. The bottom-up method is just the opposite. It starts with entities, summarizes and organizes entities to form low-level concepts, and then gradually abstracts upward to form upper-level concepts.

In the UHV projects knowledge domain, heterogeneous data comes from different sources. On the one hand, existing data models such as traditional power system engineering systems and expert knowledge bases can provide equipment information and other constraints. On the other hand, rich knowledge is buried in various types of data such as unstructured, semi-structured, and structured data, which needs to be added to the knowledge base using data mining methods. To ensure the integrity of UHV knowledge graph ontology, a combination of top-down and bottom-up is used. Expert experiences are regarded as the guidelines of the construction, while

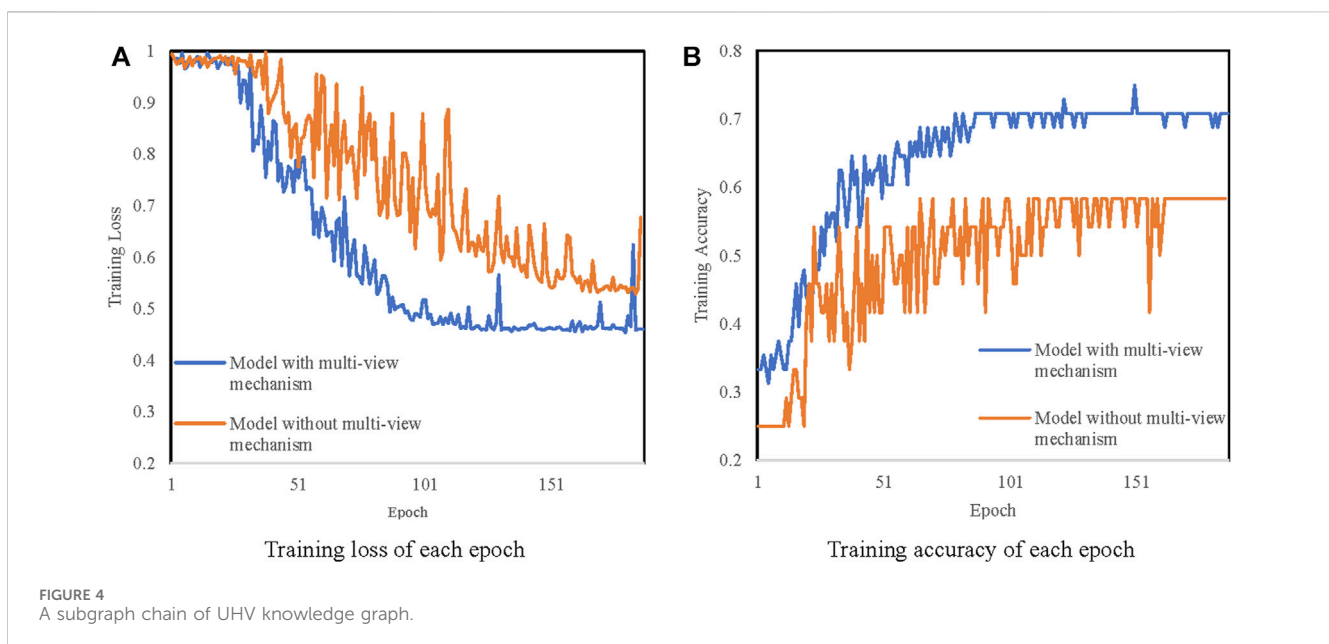
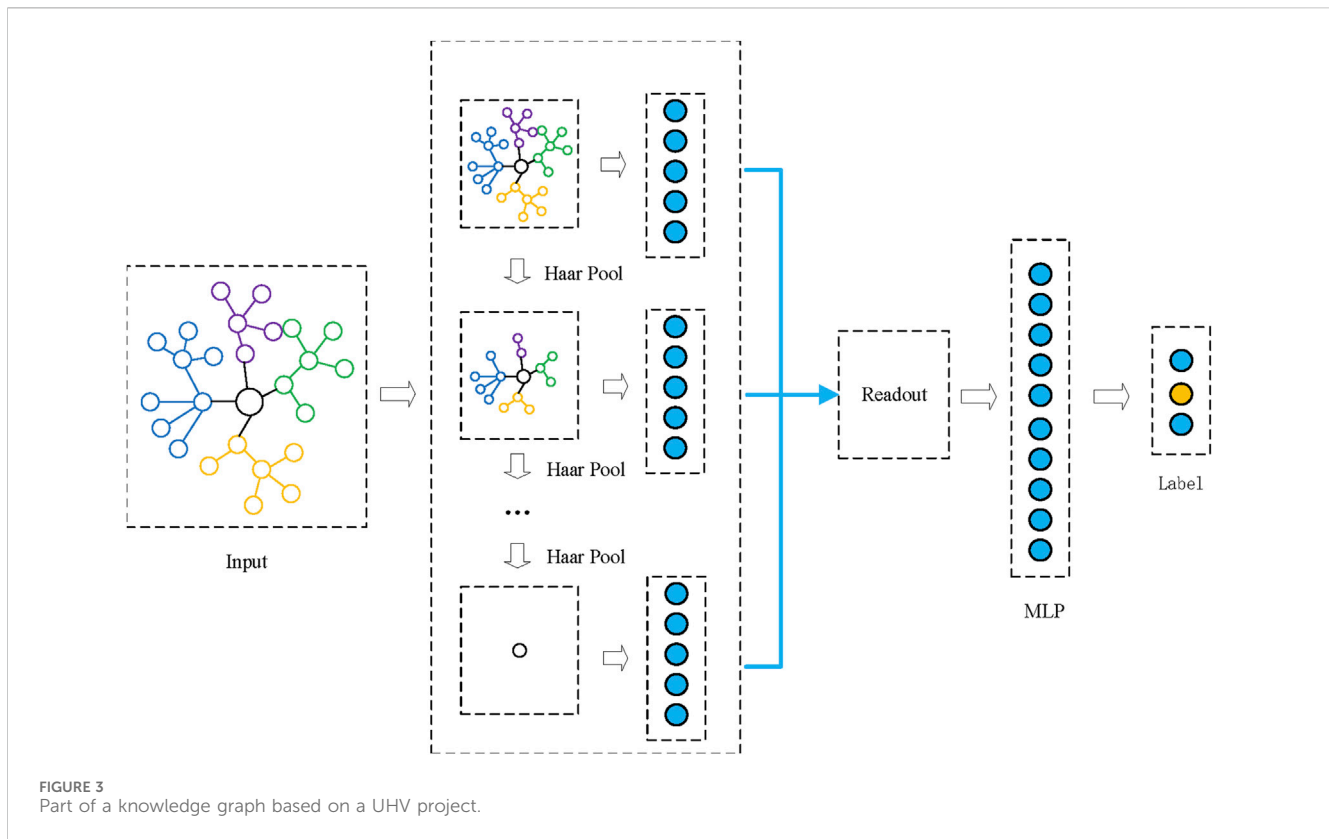
knowledge extracted from data is a supplement to the ontology. These two aspects together form the conceptual part of the knowledge graph.

2) Knowledge graph construction for UHV projects

Knowledge construction for UHV projects constitutes the data layer of the knowledge graph. It mainly contains 2 components: knowledge extraction and knowledge fusion.

Knowledge extraction is a technology that automatically extracts structured information such as entities, relationships, and entity attributes from heterogeneous data. Conventional named entity recognition (NER) algorithm includes BiBERT-LSTM-CRF (Huang et al., 2015), RoBERTa-CRF (Liu et al., 2019), and other deep neural network models. To solve the nested entity problem, methods based on entity matrix such as GlobalPointer are proposed. Other methods such as TPLinker, Tencent Multi-head, and Deep Biaffine are also suitable for solving nested NER problems. As for relation extraction, recent research mainly focuses on joint extraction methods, which can be divided into two categories: sequence annotation-based methods which converts joint extraction problem into sequence annotation problem when decoding, and sequence-to-sequence based methods (Zeng et al., 2018; Takanobu et al., 2019). For constructing a UHV knowledge graph, due to the fact that there exists a nested NER phenomenon, a combination of manual annotation and GlobalPointer is recommended. For example, given a sentence “The Beijing 1000 kV substation project requires four new main transformers”, which describes the quantity of transformers in the project. To recognize the nested entity “Beijing 1000 kV substation project”, GlobalPointer first lists all the entity candidates and then Pick out the real entities with entity labels. It can be concluded that a sentence with n words could generate $\frac{n(n+1)}{2}$ candidates. With m labels, the NER problem is then converted to a multi-label classification problem. Although the time complexity is $\sigma(n^2)$, with the use of Transformer and the design of parallel computing, it can be reduced to $\sigma(1)$.

Knowledge fusion refers to the fusion of description information about the same entity or concept from multiple data sources and the integration and disambiguation of heterogeneous data under unified standards for knowledge graphs. It requires two processes such as implement entity linking and knowledge merging. The process of entity linking is to use a given entity referent to



perform entity disambiguation and coreference resolution through similarity calculation. After confirming the correct entity object, the entity referent is linked to the corresponding entity in the knowledge graph. Among them, entity disambiguation solves the problem of ambiguity of entities with the same name, and coreference resolution solves the problem of multiple references corresponding to the same entity object.

Based on the data extracted from a UHV project, Figure 3 shows the constructed knowledge graph. As can be seen, a UHV project mainly contains six parts: electrical primary part, electrical secondary part, hydraulic part, HVAC part, remote control part and protection part, which is represented as orange nodes in the figure. Each part contains multiple equipment, with components followed.

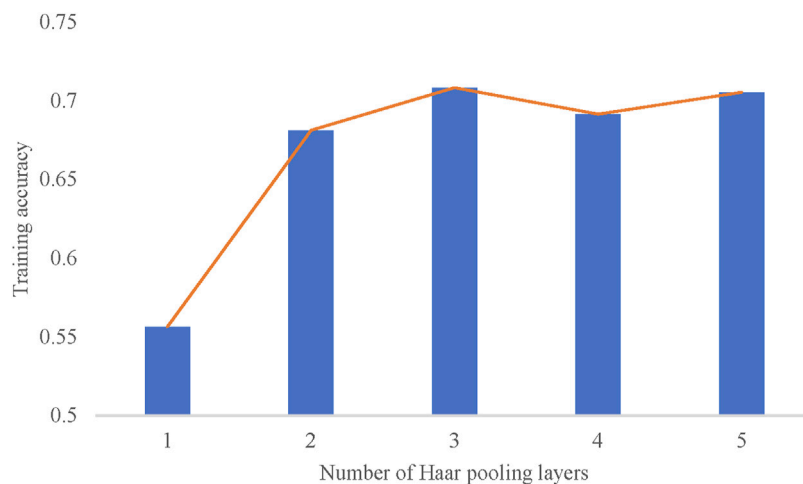


FIGURE 5
A framework of graph classification based on the Haar pooling mechanism.

3.2 Multi-view graphs encoder

After obtaining the knowledge graph of UHV projects, we use attributed network model to analyze the knowledge reasoning problem. To transform knowledge graphs to attributed networks, we take entities as nodes and edges between entities become edges between nodes. And attributes of each node are embedded before concatenated as a vector. For a specific UHV graph, we divide it into multi-view graphs considering the difference of multiple attributes of a node. Taking the accuracy and effectiveness of the multi-view divide into account, there are two principles: 1) attributes with different units need to be divided into different views in order to avoid confusion of attribute information due to excessive numerical differences; 2) pure numeric properties and mixed text and numeric properties should be separated into separate views.

After the construction of multi-view graphs, a GCN-based encoder is adopted to get the latent representation of each graph. GCN is a convolutional neural network that can directly act on graphs and utilize their attribute and structural information. To get the most precious embedding of the graph, several layers of GCN are usually used. Given a graph $G(V, A, X)$ with k views, each of the multi-view graph can be denoted as $G_i^{mg}(V, A, X_i^{mg})$, where $X = (X_1^{mg}, X_2^{mg}, \dots, X_k^{mg})$. The embedding of each layer can be denoted as:

$$\mathbf{H}^{l+1} = f(\mathbf{H}^l, \mathbf{A}, \mathbf{W}^l) \quad (2)$$

where \mathbf{H}^l is the input of the layer l , and \mathbf{W}^l is the trainable parameters.

More specifically, the function $f(\mathbf{H}^l, \mathbf{A}, \mathbf{W}^l)$ can be expressed as follows:

$$f(\mathbf{H}^l, \mathbf{A}, \mathbf{W}^l) = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^l \mathbf{W}^l\right) \quad (3)$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, $\tilde{\mathbf{D}}$ is the degree matrix of the nodes with $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, σ is an activation function such as *Relu*. To simplify the calculation, \mathbf{W}^l is shared for all the nodes. It is worth noticing that GCN can only capture the information of the neighborhood for a specific node. In order to take long-range information into consideration, a deep GCN network with k convolutional layers is

required. However due to an over-smoothing problem, the number of layers is no more than three. In this paper, a UHV knowledge graph is built with a depth of three, thus a three-layer GCN network is used. The final feature representation can be expressed as:

$$\mathbf{Z}_i^{mg} = \text{Relu}(\text{Relu}(\text{Relu}(\mathbf{X}^{mg}, \mathbf{A}, \mathbf{W}^0), \mathbf{A}, \mathbf{W}^1) \mathbf{A}, \mathbf{W}^2) \quad (4)$$

3.3 Multi-view graphs aggregator

For every low-dimensional feature representation, it is important to understand that each view only contains part of the attribute information of the nodes. To be more specific, in the UHV knowledge graph, \mathbf{Z}_1^{mg} , \mathbf{Z}_2^{mg} and \mathbf{Z}_3^{mg} represent the embedding of rated voltage, quantity and device model respectively. To get the full description of the nodes, an aggregator gathering all information from every view seems sensible. There are usually two methods when it comes to representation aggregation, concatenation and weighted addition. As each attribute is equally important to a specific node, we apply concatenation to polymerize each representation. The unified representation of the original knowledge graph can be denoted as follows:

$$\mathbf{Z} = \text{concat}(\mathbf{Z}_1^{mg}, \mathbf{Z}_2^{mg}, \mathbf{Z}_3^{mg}) \quad (5)$$

3.4 Construction of subgraphs based on Haar pooling

After obtaining the unified representation of the UHV knowledge graph, traditional algorithms take it as an input to a multilayer perceptron (MLP) and output a possibility score of each category. Because the electric power grid infrastructure domain has a strong hierarchical relationship, especially in the equipment selection area, we can use a clustering method to get the node aggregation features in the knowledge graph. Further

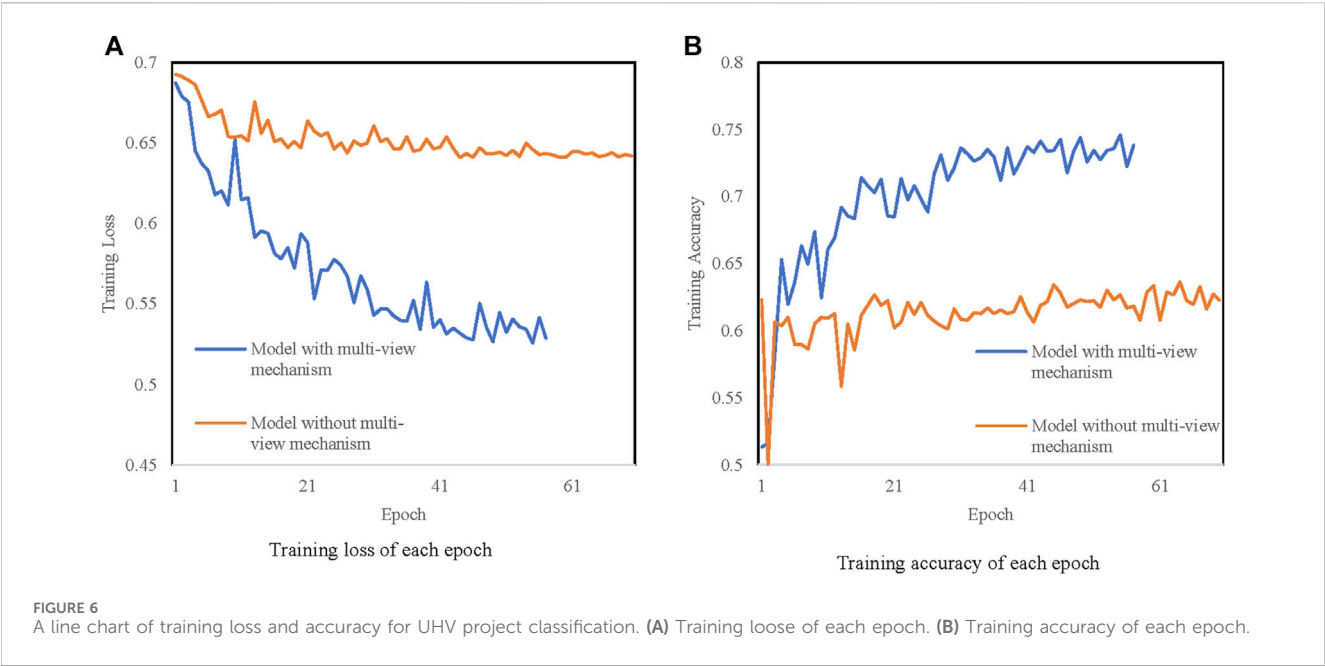
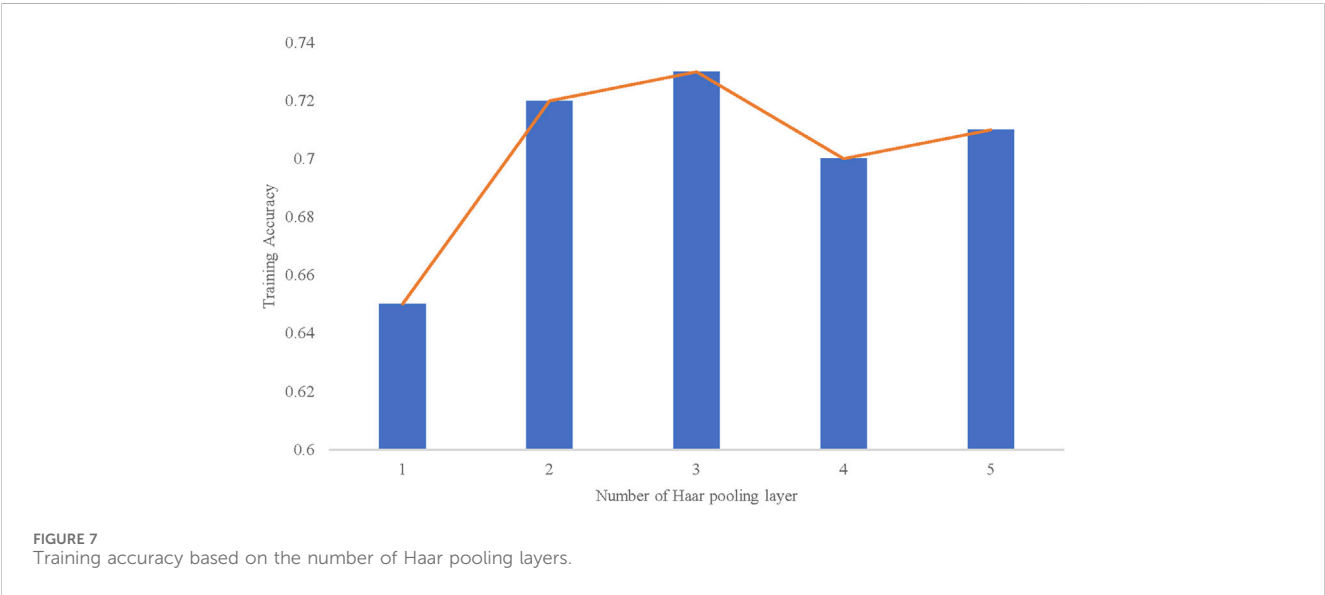


TABLE 1 UHV projects classification considering multi-view mechanism.

Algorithm	Epochs	Test accuracy	Test loss	Time
Haar pool with multi-view mechanism	188	0.7083	0.4600	0.2645
Haar pool without multi-view mechanism	187	0.5833	0.6774	0.2436

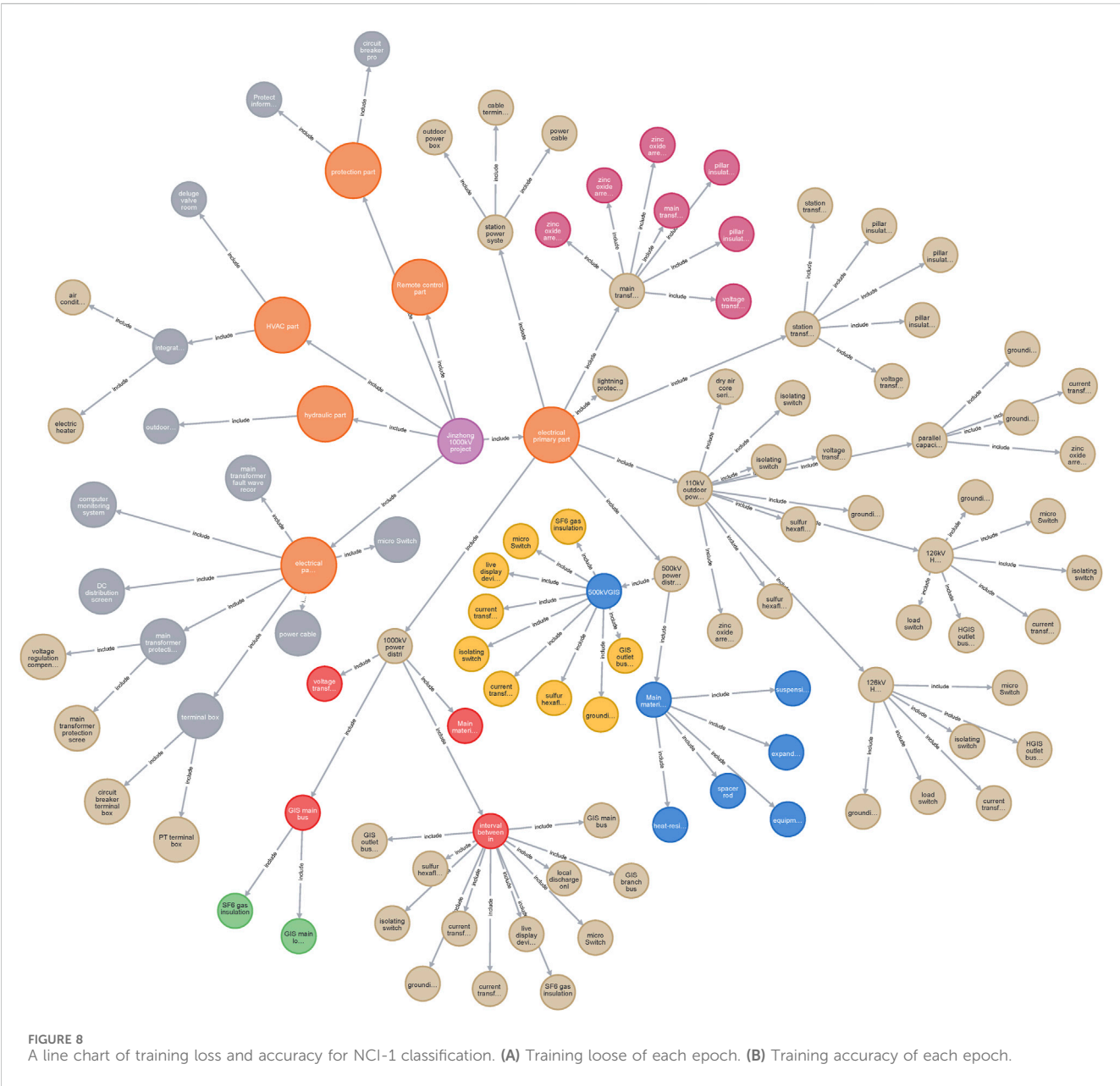


analysis of the knowledge graph shows a sparse characteristic which means the number of edges is far less than the number of nodes. Traditional algorithms such as the k-nearest neighbors algorithm(KNN) use the distance between nodes as the basis for classification, which is not suitable for graph data because it ignores the topology information. Therefore, considering the clear hierarchy relationship in components, parts, and

equipment in the electric power equipment knowledge graph, we can obtain a chain of subgraphs based on the facts of the business situation without adopting a clustering algorithm. In this context, getting a series of representation of subgraphs effectively becomes the focus of this problem. In this paper, a Haar convolution-based pooling mechanism is leveraged to the graph unified representation.

TABLE 2 Test results of different pooling mechanisms.

Algorithm	Test accuracy	Test loss	Time
Multi-view mechanism with Haar pool	0.7083	0.4600	0.2645
Multi-view mechanism with SAGPoll	0.6753	0.7521	0.2015
Multi-view mechanism with CGIPool	0.6861	0.8652	0.3054
Multi-view mechanism with GSAPool	0.6675	0.8534	0.2465



Given a chain of the subgraphs (G_0, G_1, \dots, G_m) , each node in G_i is a set of nodes in G_{i+1} . As shown in Figure 4, G_0 represents the original knowledge graph, while G_3 can be a single node that represent the whole knowledge graph. The Haar convolutional-based pooling mechanism should be used in every layer of the

chain, getting a smaller graph from G_i to G_{i+1} . Since a readout module is used afterward to integrate every representation of subgraphs, it is important to guarantee each matrix has the same output dimension. The output of a specific subgraph $G_i(V^i, E^i)$ is defined as:

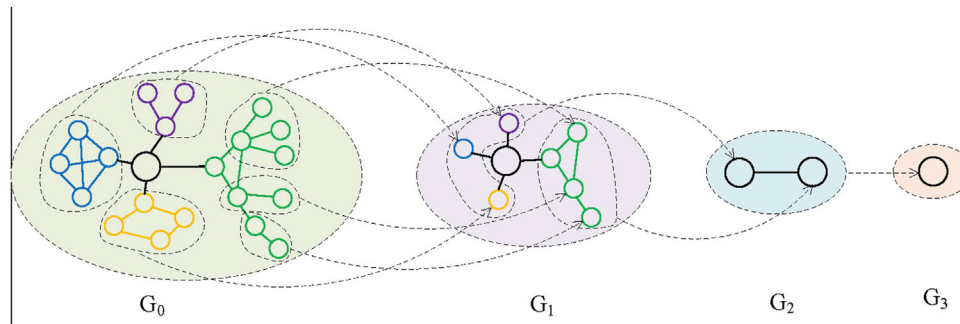


FIGURE 9
Training accuracy based on the number of Haar pooling layers.

$$X_i^{\text{output}} = \Phi_i^T X_i^{\text{input}} \quad (6)$$

$$Y = \text{softmax}(\text{MLP}(X_{\text{Readout}})) \quad (9)$$

where Φ_i^T is the transpose of Harr basis matrix of the i th subgraph, V^i , E^i represents the nodes and edges of G_i respectively.

Then the Haar basis could be obtained using the following steps: 1) order V^i by their degrees of nodes as $\{v_1^i, v_2^i, \dots, v_N^i\}$, where N denotes the number of nodes in G_i ; 2) the N vectors can be calculated by the following equations:

$$\Phi_l^i = \frac{1}{\sqrt{N}} \quad l = 1, \quad (7)$$

$$\Phi_l^i = \sqrt{\frac{N-l+1}{N-l+2}} \left(x_{l-1}^i - \frac{\sum_{j=l}^N x_j^i}{N-l+1} \right)$$

where x_j^i is a function for j th node, which could be expressed as:

$$x_j^i(v) = \begin{cases} 1, v = v_j^i \\ 0, v \in V^i \setminus \{v_j^i\} \end{cases} \quad (8)$$

To speed up the calculation process, a compressed Haar basis matrix is used to replace the original Φ_i . Considering the fact that the first N_{i+1} represents the low frequency coefficients which contains the majority information of the graph, we adopt a tailor operation to reduce the size of Φ_i . The Haar convolution-based pooling mechanism requires less computational complexity than the traditional graph Fourier transform. The use of fast algorithms further improves calculation speed. The key to the fast algorithms is the Haar basis, which are sparse matrixes to which a compressive method can be adopted. Specifically, given a chain of subgraphs (G_0, G_1, \dots, G_m) , where G_m is the coarsest graph with only one node. Each subgraph could generate a set of n orthogonal vectors called Haar basis, where n is the number of nodes in the specific subgraph. They are the representations of the input in the Haar domain. Analysis shows that these matrixes are sparse and several low-frequency coefficients contains most of the information of the input. Therefore, we use first n_{j+1} Haar basis vector in our pooling layer, while the other components are aborted.

In order to minimize errors caused by information loss, a hierarchical mechanism is adopted. As shown in Figure 5, every low-dimensional representation is aggregated together using a Readout module and then an MLP is used to get the probability of each label for subgraph G_i .

4 Case studies

In this section, we tested our proposed method on two datasets along with several traditional methods as comparison algorithms. The two datasets are UHV_Projects and NCI-1. Among them, UHV_Projects is constructed on historical engineering data, and NCI-1 is an open-source dataset widely used to test the classification performance of different algorithms. All these experiments are performed on a laptop with Intel Core i7-11800H CPU and NVIDIA GeForce RTX 3070 Laptop GPU. Some of the hyperparameters are set as follows: batch size is fixed to 50; learning rate is fixed to 0.001.

4.1 UHV projects data classification

The construction of the UHV projects knowledge graph contains heterogeneous data from multiple data sources such as Preliminary Design Instructions, Equipment Inventory, Specialized Reports, and so on. It requires efforts from experts from both the power electrical domain and computer domain. Due to the complexity of different business needs, we only construct a dataset based on the requirements of power transmission and transformation project review with a focus on equipment. Therefore, the dataset mainly contains four levels: the first level is the name of the projects; the second level contains six professional fields based on construction guidelines, such as electrical primary part, electrical secondary part, protection part, telecontrol part, hydraulic part and fire-controlling part; the third level mainly refers to major equipment of each part of second level, such as main transformers, 1000 kV power distribution unit; the tertiary level contains parts of the major equipment, such as cable, voltage transformer, current transformer, breaker and so on.

Taking each project as a subgraph, a label is assigned to each graph indicating the type of the project, which are the new substation project, substation expansion project, and bay expansion project. The difference between the three kinds of projects mainly lies on the quantity and equipment models. For

TABLE 3 NCI-1 classification considering multi-view mechanism.

Algorithm	Epochs	Test accuracy	Test loss	Time
Haar pool with multi-view mechanism	57	0.7385	0.5286	5.5319
Haar pool without multi-view mechanism	70	0.6223	0.6419	4.8624

TABLE 4 Test results of different pooling mechanisms.

Algorithm	Test accuracy	Test loss	Time
Multi-view mechanism with Haar pool	0.7385	0.5286	5.5319
Multi-view mechanism with SAGPoll	0.6967	0.6519	5.1864
Multi-view mechanism with CGIPool	0.7058	0.7622	5.9571
Multi-view mechanism with GSAPool	0.6885	0.6681	5.3748

example, a new substation project usually has more main transformers than substation expansion projects and a bay expansion project may only have some in-bay equipment such as isolating switches and lightning protectors. Thus, we constructed 120 graphs, with an average of 130 nodes and 128 edges per graph. We first test our model and our model without multi-view mechanism on this dataset and the results are illustrated in Figure 6 and Table 1. It can be seen that the accuracy of our model with a multi-view mechanism is improved by 21.43% with a little more time consumption.

We also add an analysis regarding the number of Haar pooling layers. It describes the number of subgraphs which is an indicator of different clustering methods. It should be noted that 1 layer represents the model without the Haar pooling strategy, and 2 layers only consist of two subgraphs which are the original knowledge graph and the final knowledge graph with only one node. The results are shown in Figure 7. Due to the fact that the UHV knowledge graph could be divided into three levels which are professional domains, equipment, and parts, a three-level clustering is most proper and the result of the experience also proves the conclusion.

To test the performance of the Haar pool, we selected three traditional pooling methods as the control group such as SAGPoll (Lee et al., 2019), CGIPool (Pang et al., 2021), and GSAPool (Zhang et al., 2020). The results are illustrated in Table 2. According to test results of different pooling mechanisms, our proposed model shows excellent performance both in accuracy and loss. Although a multi-view mechanism with SAGPoll achieves the highest computational efficiency, the accuracy and the loss of our proposed model outperforms it by 4.88% and 63.5%. It shows that a little sacrifice of the time complexity returns good performance.

4.2 NCI-1 data classification

To test the generalizability of our proposed model, we use the NCI-1 dataset as our input. NCI-1 is a popular open-source benchmark dataset mainly focused on chemical and medical domain. It contains 4,100 compounds, each one of which could be seen as a graph sample. Among them, nodes represent atoms and edges represent chemical bonds. The task

of the dataset is to determine whether the compound has properties that hinder the growth of cancer cells. As experiments in Section 4.1, we first test multi-view mechanism on this dataset. The results are shown in Figure 8 and Table 3. It shows an increase in test accuracy by 18.67% when comparing the model with a multi-view mechanism to the one without. As the number of samples increases, the running time is also extended. The difference between both algorithms is extended when compared to UHV datasets, which implies that traditional GCN has a non-linear computational complexity.

As for the number of Haar pooling layers, there is no obvious structure of the NCI-1 dataset. We make a test from layer 1 to layer 5. The results are shown in Figure 9. According to training accuracy, apart from the results for 1 layer, the rest of the results are almost the same.

To test the performance of the Haar pool, we also employ the traditional pooling mechanism as in Section 4.1. And the results are shown in Table 4. It demonstrates that our proposed method is best-considering test accuracy and test loss. Due to the multi-view mechanism, the running time is not the shortest of all experiments, but considering the increase in performance, it is completely acceptable.

5 Conclusion

This paper proposed a novel multi-view knowledge reasoning method that takes both attribute and structural characteristics into consideration. Firstly, a knowledge graph construction method is proposed based on UHV project data. Secondly, considering the difference of multiple attributes, a series of multi-view graphs are constructed and represented using traditional GCN. Thirdly, a Haar convolutional-based pooling method is leveraged to deal with the structural information with high efficiency. Results from the UHV dataset and NCI-1 dataset prove the feasibility of our algorithm. In general, our contributions are as follows:

- 1) The introduction of a multi-view mechanism to the knowledge reasoning framework improves the accuracy of graph representation learning.

- 2) A Haar convolutional-based pooling mechanism is used in the UHV knowledge graph, which proposes a better way when analyzing knowledge graphs with hierarchical structures.
- 3) Although running time increases, our proposed method shows various improvements in accuracy in different datasets.

In the future, knowledge graphs will become more and more popular in the electric power domain. Knowledge reasoning methods should surely serve in various downstream applications. Our proposed method mentioned above mainly focus on graph classification and could be used in power transmission and transformation review platform.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

JH: Conceptualization, Writing–original draft. GX: Supervision, Writing–review and editing. LQ: Supervision, Writing–review and editing. XQ: Supervision, Writing–review and editing.

References

- Gao, F., Li, G., Gu, J., Zhang, L., and Wang, L. (2023). GridOnto: knowledge representation and extraction for fault events in power grid. *IEEE Access* 11, 58863–58878. doi:10.1109/ACCESS.2023.3284839
- Hamilton, W., Ying, Z., and Leskovec, J. (2017). Inductive representation learning on large graphs. *Adv. neural Inf. Process. Syst.* 30. doi:10.48550/arXiv.1706.02216
- Huang, J. (2018). Research on intelligent review management platform for 110 kV and below power transmission and transformation project design. *M. Electr. Tech.* 41, 112–115. doi:10.16652/j.issn.1004-373x.2018.04.028
- Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. Available at: <https://arxiv.org/abs/1508.01991>.
- Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S. (2022). A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 494–514. doi:10.1109/TNNLS.2021.3070843
- Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. Available at: <https://arxiv.org/abs/1609.02907>.
- Lee, J., Lee, I., and Kang, J. (2019). “Self-attention graph pooling,” in Proceedings of the International conference on machine learning, Long Beach, United States, June 2019, 3734–3743.
- Li, G., Zhou, X., and Wang, B. (2021). Research on intelligent review and management platform for power transmission and transformation engineering design of 110kV and above. *Cn. M. Inf.* 24, 107–108.
- Liu, J., Tang, Q., Su, Y., Li, T., Wang, Y., and Zhu, M. (2021). “Economic analysis of solid oxide fuel cell and its role in carbon peak, carbon neutralization process,” in Proceedings of the 2021 4th International Conference on Energy, Electrical and Power Engineering (CEEPE), Chongqing, China, April 2021, 115–119.
- Liu, L., and Yan, X. (2023). Knowledge graph in distribution network fault handling: advances, challenges and prospects. *E. P. Info. Comm. Technol.* 21, 19–26. doi:10.16543/j.2095-641x.electric.power.ict.2023.07.03
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). Roberta: a robustly optimized bert pretraining approach. Available at: <https://arxiv.org/abs/1907.11692>.
- Luo, Q., Mao, Z., Zhao, Q., Gong, S., Zhao, H., You, H., et al. (2023). “China’s energy evolving approach to carbon peaking by 2030 based on 23 provincial carbon peaking implementation programs,” in Proceedings of 2023 Panda Forum on Power and Energy (PandaFPE), Chengdu, China, April 2023, 1076–1080.
- Pang, Y., Zhao, Y., and Li, D. (2021). “Graph pooling via coarsened graph infomax,” in Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA, July 2021, 2177–2181.
- Peng, Z., Luo, M., Li, J., Xue, L., and Zheng, Q. (2020). A deep multi-view framework for anomaly detection on attributed networks. *IEEE Trans. Knowl. Data Eng.* 34, 1–2552. doi:10.1109/TKDE.2020.3015098
- Pu, T., Tan, Y., Zheng, G., Xu, H., and Zhang, Z. (2021). Construction and application of knowledge graph in the electric power field. *P. Sys. Technol.* 45, 2080–2091. doi:10.13335/j.1000-3673.pst.2020.2145
- Ren, Y., Ma, C., Chen, H., and Huang, J. (2021). “Low-carbon power dispatch model under the carbon peak target,” in Proceedings of the 2021 IEEE Sustainable Power and Energy Conference (iSPEC), Nanjing, China, December 2021, 1–5.
- Takanobu, R., Zhang, T., Liu, J., and Huang, M. (2019). “A hierarchical framework for relation extraction with reinforcement learning,” in Proceedings of the AAAI conference on artificial intelligence, Nanjing, China, January 2019, 7072–7079.
- Tian, J., Song, H., Chen, L., Sheng, G., and Jiang, X. (2022). Entity recognition approach of equipment failure text for knowledge graph construction. *P. Sys. Technol.* 46, 3913–3922. doi:10.13335/j.1000-3673.pst.2021.1886
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. Available at: <https://arxiv.org/abs/1710.10903>.
- Wu, X., Jiang, T., Zhu, Y., and Bu, C. (2023). Knowledge graph for China’s genealogy11.a shorter version of this paper won the best paper award at icke 2020 (the 11th icke international conference on knowledge graph, ickg 2020.bigke.org). *IEEE Trans. Knowl. Data Eng.* 35, 634–646. doi:10.1109/TKDE.2021.3073745
- Yang, J., Yang, L. T., Wang, H., Gao, Y., Liu, H., and Xie, X. (2022). Tensor graph attention network for knowledge reasoning in internet of things. *IEEE Internet Things J.* 9, 9128–9137. doi:10.1109/JIOT.2021.3092360
- Zeng, X., Zeng, D., He, S., Liu, K., and Zhao, J. (2018). “Extracting relational facts by an end-to-end neural model with copy mechanism,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, July 2018, 506–514.
- Zhang, L., Wang, X., Li, H., Zhu, G., Shen, P., Li, P., et al. (2020). “Structure-feature based graph self-adaptive pooling,” in Proceedings of The Web Conference 2020, Taipei, China, April 2020, 3098–3104.
- Zou, Q., and Lu, C. (2022). An accurate matching query method of natural language knowledge graph based on hierarchical graph topological sequence. *IEEE Access* 10, 24080–24094. doi:10.1109/ACCESS.2022.3155520

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

Author LQ was employed by State Grid Economic and Technological Research Institute Co Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



OPEN ACCESS

EDITED BY

Yikui Liu,
Stevens Institute of Technology, United States

REVIEWED BY

Ge Cao,
Xi'an University of Technology, China
Wei Gao,
University of Denver, United States
Siyuan Chen,
Wuhan University, China

*CORRESPONDENCE

Ziyong Ye,
✉ zjlszy97@163.com

RECEIVED 17 September 2023

ACCEPTED 08 January 2024

PUBLISHED 19 January 2024

CITATION

Zou H, Ye Z, Sun J, Chen J, Yang Q and Chai Y (2024), Research on detection of transmission line corridor external force object containing random feature targets. *Front. Energy Res.* 12:1295830. doi: 10.3389/fenrg.2024.1295830

COPYRIGHT

© 2024 Zou, Ye, Sun, Chen, Yang and Chai. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Research on detection of transmission line corridor external force object containing random feature targets

Hongbo Zou^{1,2}, Ziyong Ye^{2*}, Jialun Sun², Junting Chen²,
Qinhe Yang² and Yanhui Chai²

¹Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station and New Energy, China Three Gorges University, Yichang, China, ²College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, China

With the objective of achieving “double carbon,” the power grid is placing greater importance on the security of transmission lines. The transmission line corridor has complex situations with external force targets and irregularly featured objects including smoke. For this reason, in this paper, the high-performance YOLOX-S model is selected for transmission line corridor external force object detection and improved to enhance model multi-object detection capability and irregular feature extraction capability. Firstly, to enhance the perception capability of external force objects in complex environment, we improve the feature output capability by adding the global context block after the output of the backbone. Then, we integrate convolutional block attention module into the feature fusion operation to enhance the recognition of objects with random features, among the external force targets by incorporating attention mechanism. Finally, we utilize EIoU to enhance the accuracy of object detection boxes, enabling the successful detection of external force targets in transmission line corridors. Through training and validating the model with the established external force dataset, the improved model demonstrates the capability to successfully detect external force objects and achieves favorable results in multi-class target detection. While there is improvement in the detection capability of external force objects with random features, the results indicate the need to enhance smoke recognition, particularly in further distinguishing targets between smoke and fog.

KEYWORDS

transmission line corridor, external force, object detection, random feature targets, attention mechanism

1 Introduction

China is undergoing rapid modernization, and one of the fundamental aspects of this process is the expansion of the power transmission network. The electric power industry, as a crucial component of the national infrastructure, plays a significant role in the development of the country's economy. China has established explicit goals of achieving “carbon peaking” by 2030 and “carbon neutrality” by 2060. Consequently, the country will prioritize promoting adjustments in its industrial and energy structure. Additionally, the power grid will increasingly integrate a significant share of clean energy sources, such as wind and photovoltaic power (Hu, et al., 2022; Tian et al., 2022; Xiao and Zheng, 2022). As a result, there will be an increased need for secure operation of the transmission line

corridor. These transmission line corridors connect power production sources to the load, necessitating their secure operation. Given the vastness of China, variations in climate, and other environmental conditions, the establishment of transmission line corridor poses unique challenges. Consequently, the security of the transmission line corridor is frequently compromised. The destruction of transmission line is caused not only by fires but also by various other factors (Gu, et al., 2020; Sheng et al., 2021). Additionally, the presence of large tower cranes and engineering machinery also poses a significant risk. Statistical data indicates that the primary causes of transmission line tripping, in order, are lightning strikes, external force damage, wind deviation, and ice damage, among others. Among these causes, external force damage accounts for 21.4% of transmission line trips, second only to those caused by lightning strikes. Furthermore, an analysis of faults in transmission lines of 220 kV and above in the provincial power grid revealed that external force damage accounted for 12.36% of all transmission line tripping, resulting in a low success rate of tripping and reclosing at only 44.7% (Liang, 2014; Lu, et al., 2016). Additionally, the outage durations typically range between 2 and 3 h. It can lead to damage to the grid infrastructure, which can disrupt production, leading to avoidable economic losses. Additionally, it poses a risk to the safety of civilians, staff, and individuals in close proximity to the electrical equipment.

In current method for ensuring safety, manual inspection is predominantly utilized as the primary method for preventing external force damage. Nonetheless, this method is time-consuming, and the constraint of limited number of workers makes it challenging to continuous inspection (Wang, et al., 2019; Wang, et al., 2021; Ma, et al., 2022). The researcher conducted monitoring of external force in transmission lines using helicopter and drone video surveillance (Golightly and Jones, 2003; Larrauri et al., 2013; Lin, et al., 2019; Wei, et al., 2022). This method has the potential to enhance monitoring efficiency and reduce labor requirements. However, it is important to note that the patrol monitoring method is unable to provide real-time monitoring of moving targets within a specified area. Hence, it is not applicable for detecting unauthorized construction machinery operating within transmission line corridors. The reference (Zhang and Deng, 2020) obtains the external force vibration signal of transmission pole and the vibration signals of transmission towers under different wind excitation conditions. The vibration signals are preprocessed by adopting delay inlay technology to turn the original signal into a two-dimensional form and sent that into convolutional neural network to feature extraction, and achieve vibration pattern recognition by employ the relevance vector machine (Cui, et al., 2023). However, this method has limitations when it comes to recognizing different types of external force.

Detecting external force in transmission line corridor through image analysis enables the reduction of manual labor, effectively alerting against potential threats and preventing external force incidents. The continuous development and application of deep learning in the field of image recognition has proven invaluable for identifying external force damages in transmission lines (Krizhevsky, et al., 2012; Redmon, et al., 2016; Zhang, et al., 2018; Liu, et al., 2019; Ma, et al., 2021; Long, et al., 2022; Wu, et al., 2022; Dong, et al., 2023). The reference (Wei, et al., 2021) uses bounding box annotation instead of partial mask annotation in the process of dataset annotation and improves the average accuracy of

recognizing common categories of external force. The reference (Tian, et al., 2021) employs an enhanced K-means algorithm to determine suitable anchor box from the image. Subsequently, the CSP Darknet-53 residual network is used to extract the deep-seated network feature data of the images, and the feature map is processed by the SPP algorithm. The algorithm, when applied to real-time monitoring pictures of transmission line, demonstrates its ability to detect external force damage accurately and promptly.

Nowadays, the monitoring devices can collect and transfer image. The devices equipped with target recognition model enable target detection in real-time images. Nevertheless, most image monitoring systems solely recognize targets exhibiting distinct features, such as prominent engineering machinery and construction scenes. They often fail to identify targets with irregular characteristics like smoke. In this study, we enhance the performance of the YOLOX-S model by incorporating the global context block (GC block) to enhance the feature output capacity of the backbone network. Additionally, we introduce the convolutional block attention module (CBAM) in the feature fusion process to improve the recognition of randomly featured targets, among externally damaged objects. Finally, we employ the EIoU to enhance the accuracy of target detection box and achieve the detection of complex external force targets in transmission lines corridor.

2 Improved YOLOX-S

2.1 YOLOX-S

The YOLOX-S is a high-performance one-stage object detection network (Ge, et al., 2021). The network incorporates significant advancements in object detection, including decoupled heads, data augmentation, and anchor free, into the YOLO architecture. The model is composed of three main components. The backbone feature extraction network utilizes the CSP darknet architecture. The Neck enhances the feature extraction network through the use of the path-aggregation network (PANet). The prediction part employs three decoupled heads.

The backbone conducts low-level feature extraction on the input image, resulting in three feature layers. The Neck subsequently performs high-level feature extraction on these layers. Finally, three decoupled heads are employed for object detection, and the detection results are obtained accordingly.

The mosaic data augmentation algorithm is applied in the input layer. Its primary purpose is to combine four images, each accompanied by its respective box. Once the four images are spliced, a new image is generated, along with its corresponding box. Subsequently, the newly generated image is fed into the network for learning.

In addition to YOLOX-S, YOLOX has other types of networks, namely, YOLOX-Nano, YOLOX-Tiny, YOLOX-M, YOLOX-L, YOLOX-X. The performance of these models on the coco dataset is shown in Table 1. As shown in the table, YOLOX-Nano and YOLOX-Tiny are lightweight networks, so there are fewer network parameters and poorer performance at AP (%). In the rest models, the model accuracy improves as the model parameters increase, but the latency also increases. The application scenario in this study is

TABLE 1 Comparison of YOLOX in terms of AP (%) on COCO. YOLOX-Nano and YOLOX-Tiny are tested at 416 × 416 resolution, else are tested at 640 × 640 resolution.

Models	AP (%)	Parameters (M)	Latency (tested on tesla V100)
YOLOX-Nano	25.3	0.91	/
YOLOX-Tiny	32.8	5.06	/
YOLOX-S	39.6	9.0	9.8 ms
YOLOX-M	46.4	25.3	12.3 ms
YOLOX-L	50.0	54.2	14.5 ms
YOLOX-X	51.2	51.2	17.3 ms

characterized by two main features: 1) it is applied in outdoor environments, close to the edge end, and 2) the detection requires high real-time performance. For these two reasons, it is necessary to choose a base model that is moderate in model size, recognition accuracy and latency. The two lightweight models have fewer parameters and there is no dataset enhancement step in the training phase, which makes it difficult to meet the requirements for recognition accuracy. By comparing the rest of the models, YOLOX-S with fewer parameters is easier to deploy on devices used outdoors and meets the real-time requirements for recognition as well as the needs in detection accuracy.

2.2 Global context block

In transmission line corridors characterized by complex environments and a multitude of targets to be identified. The convolutional operation of the backbone network models the context within a limited range, and creating a confined receptive field. The global context block incorporates both the non-local network (NL Net) and squeeze excitation networks (SE Net). Additionally, in the GC block, the NL block is simplified to decrease the computational load. The GC block serves a dual purpose: extracting global information from the backbone convolutional network, thereby facilitating feature fusion with the linked neck part, and reducing computation cost. Therefore, incorporating the GC block into the network enables the extraction of global contextual information pertaining to external force targets in complex background (Cao, et al., 2019).

2.2.1 Simplified NL block

The NL block conducts inter-pixel correlation analysis by utilizing the current pixel in the given position along with feature-similar pixels of equal size to establish connections between features and global information. The output of the NL block z_i and the inter-pixel correlation w_{ij} are calculated using Eq. 1 and Eq. 2:

$$z_i = x_i + W_z \sum_{j=1}^{N_p} \frac{f(x_i, x_j)}{C(x)} (W_v \cdot x_j) \quad (1)$$

$$w_{ij} = \frac{f(x_i, x_j)}{C(x)} = \frac{e^{(W_q x_i)^T (W_k x_j)}}{\sum_{\forall m} e^{(W_q x_i)^T (W_k x_m)}} \quad (2)$$

Where, W_v, W_k, W_q and W_z represent the convolution operations. x_i represents the pixel at the current location, while N_p represents the set of all pixels. Additionally, x_j comprises pixels with similar features and an equal size as the location of x_i . $f(x_i, x_j)$ denotes the relationship between position i and j , and has a normalization factor $C(x)$.

The NL block computational cost in the global feature extraction process necessitates structural simplification before fusion with the GC block. As can be seen in Figure 1, the main simplifying operations include: omitting any further operations on x_i and solely considering the global pixel-feature correlation; removing the W_q convolution operation; substituting W_z with W_v , and not retaining the convolution operation of W_v . The context modeling module is composed of the simplified NL block. The mathematical expression for the simplified block is presented in Eq. 3.

$$z_i = x_i + W_v \sum_{j=1}^{N_p} \frac{e^{W_k x_j}}{e^{W_k x_m}} x_j \quad (3)$$

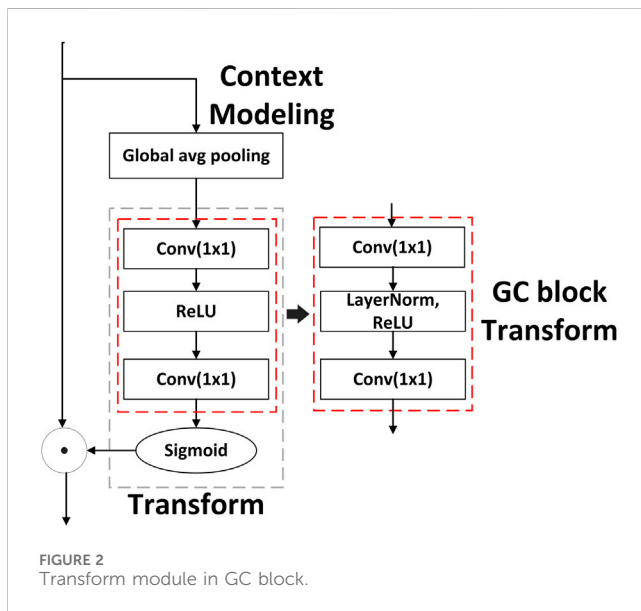
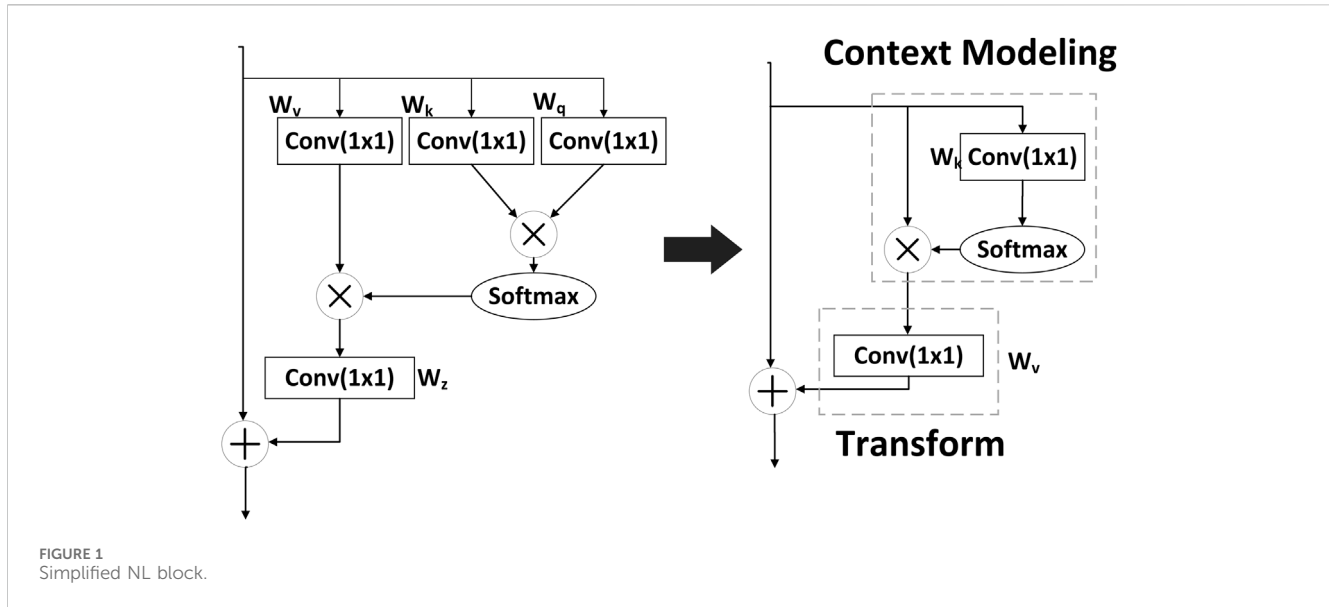
2.2.2 Transform module

Simplifying the operation of the NL block decreases the computational effort of the module but has an impact on the accuracy of the training results. To address this accuracy loss and effectively utilize the feature information from the convolution operation on the channel, the transform module within the SE block is introduced. Additionally, incorporating layer normalization before applying the ReLU nonlinear activation can enhance the generalization capability of the network. The output of transform module $z_{i,tran}$ is as Eq. 4:

$$z_{i,tran} = W_{v2} \text{ReLU} \left(\text{LN} \left(W_{v1} \left(\sum_{j=1}^{N_p} \frac{e^{W_k x_j}}{e^{W_k x_m}} x_j \right) \right) \right) \quad (4)$$

Where, W_{v1}, W_{v2} represent the convolution operations. ReLU represents the nonlinear activation function, and LN denotes layer normalization. The reconstruction of the transform module in the GC block is shown in Figure 2.

The GC block integrates the context modeling module from the simplified NL block, introduces the standardized transform module within the SE block, substitutes the W_v convolution operation in the simplified NL block, and conducts feature fusion through a sum operation. Figure 3 illustrates the final module, and its output is obtained as Eq. 5:



$$z_i = x_i + W_{v2} \text{ReLU} \left(\text{LN} \left(W_{v1} \left(\sum_{j=1}^{N_p} \frac{e^{W_k x_j}}{e^{W_k x_m}} x_j \right) \right) \right) \quad (5)$$

The GC block reduces computational parameters and computation by simplifying the NL block. The simplified block can still learn the global context. The SE block adopts rescaling to recalibrate the importance of channels but inadequately models long-range dependency. Finally, the GC block completes the feature fusion by addition to the original feature map.

2.3 Convolutional block attention module

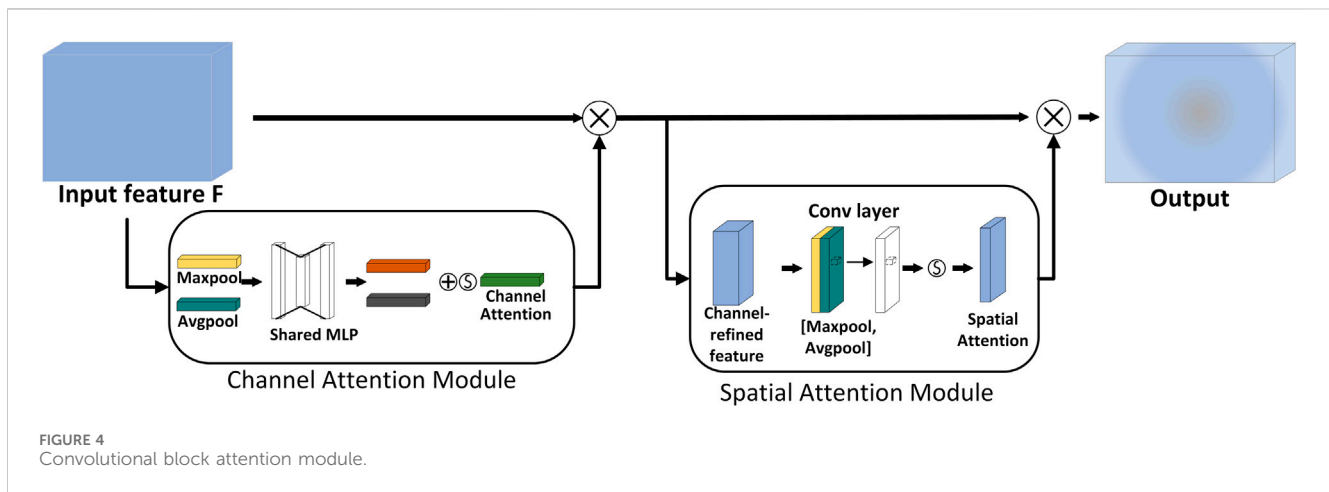
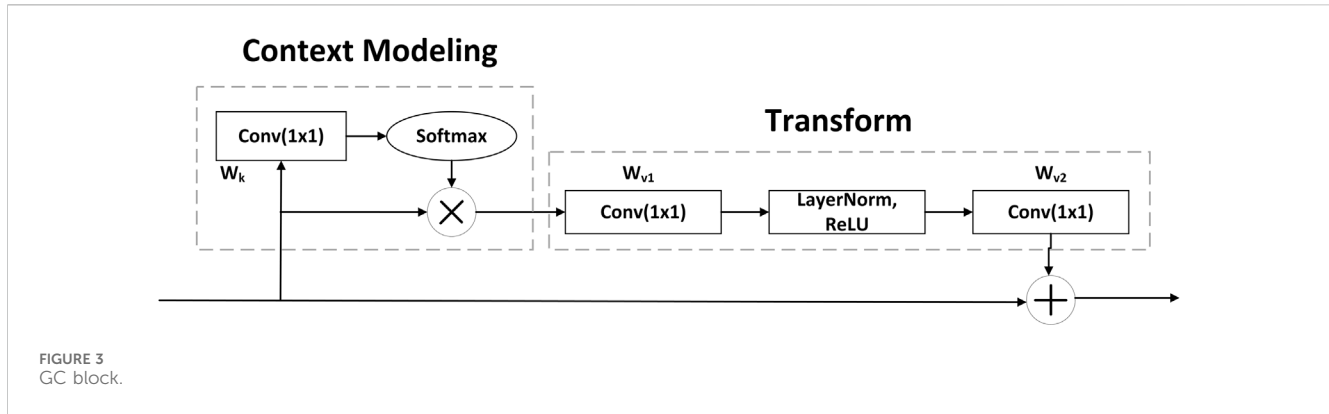
In the context of recognizing external force object in transmission lines corridor, the task involves identifying both

obvious features like engineering machinery and random features like smoke. To enhance the recognition ability for smoke, a convolutional attention mechanism is incorporated into the network. This module enables the network to simultaneously attend to multiple types of external force targets across channels and spatial dimensions during feature fusion.

The structure of the convolutional block attention module model incorporates the channel attention module (CAM) and the spatial attention module (SAM). Figure 4 depicts the convolutional block attention module (Woo, et al., 2018). Channel attention focuses on the classification of the object in the image through the channel relations of the features. Max pooling can strengthen the unique object feature, so CAM completes the object feature extraction by average pooling and max pooling. Spatial attention focuses on the localization of the target in the image. Pooling operations along the channel axis can effectively highlight information regions, so average pooling and max pooling are applied along the channel axis in SAM. CAM and SAM are placed in a sequential manner, highlighting the location of the target in the image. The sequential arrangement enables the channel and spatial attention modules to achieve complementary attention and accomplish attentional enhancement.

2.3.1 Channel attention module

The channel attention module starts by applying a global average pooling and a global max pooling method to the input's individual feature layers. The outcomes of these pooling methods are then fed into a shared fully connected layer. The results from both pooling methods are summed, and the Sigmoid activation function is applied to obtain a weighted value (from 0 to 1) for each channel in the incoming feature layer. After obtaining the weight, we multiply it by the original input feature layer to obtain the feature map processed by the channel attention module. The formula for the output of channel attention module $M_c(F)$ is as Eq. 6:



$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (6)$$

Where, F represents input feature, σ represents the Sigmoid function, F_{avg}^c and F_{max}^c refer to the feature maps obtained through average pooling and max pooling, respectively, multilayer perceptron (MLP) is the shared network in CBAM, and W_1 and W_2 represent the weights of the hidden layer and output layer in the multilayer perceptron, respectively.

2.3.2 Spatial attention module

Max pooling and average pooling are performed over the feature layer channels which come from the channel attention module. The pooling results are concatenated, and the channel number is adjusted using a convolution kernel with the size of 7×7 . After applying the Sigmoid activation function, the weights (from 0 to 1) for each feature of the input layer are obtained. The formula for the spatial attention module $M_s(F')$ and the output of CBAM F'' is as is as Eqs 7–9:

$$F' = M_c(F) \otimes F \quad (7)$$

$$M_s(F') = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (8)$$

$$F'' = M_s(F') \otimes F' \quad (9)$$

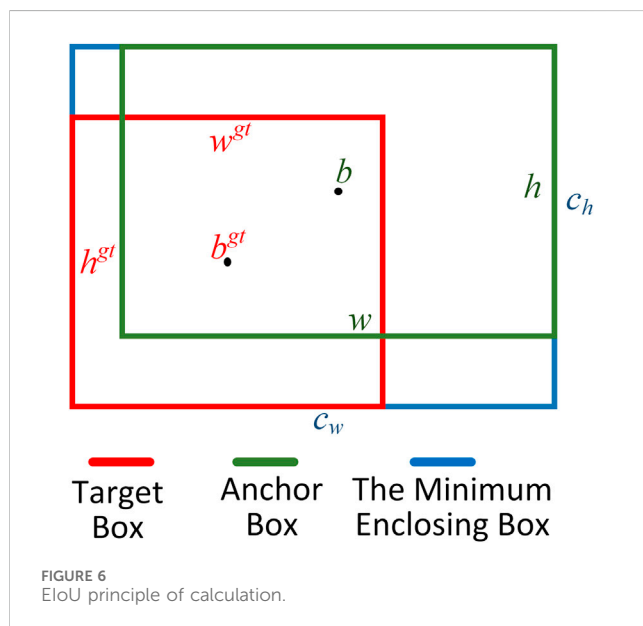
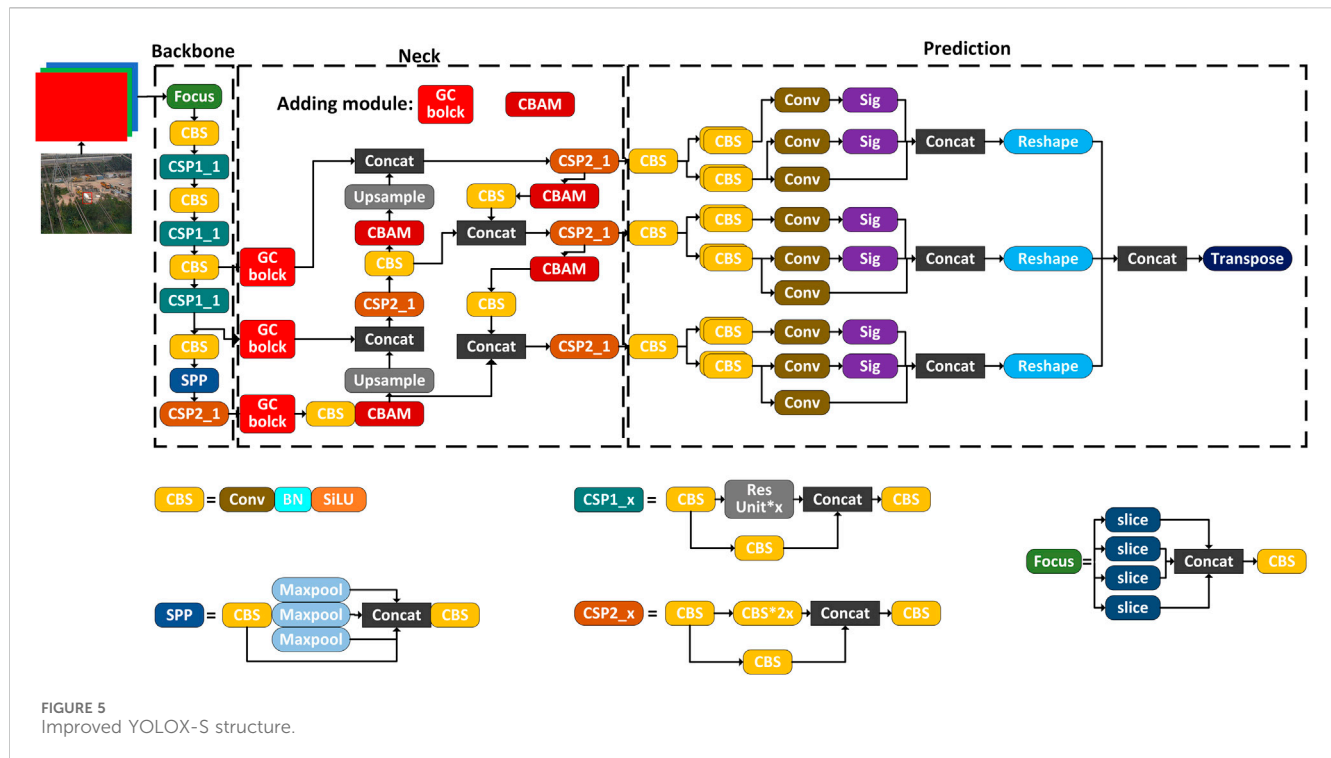
Where, F' represents channel-refined feature, σ donates the Sigmoid function, F_{avg}^s and F_{max}^s refer to the feature maps after two

pooling operations, $f^{7 \times 7}$ represents a convolution operation with the filter size of 7×7 .

In the improved YOLOX-S model, GC blocks are incorporated after the output of the backbone to augment the model's ability to perceive global features. Additionally, the CBAM is introduced to improve attention towards the external force target before the up-sampling in the feature fusion operation and before the feature contact. The modification leads to the final improved network structure of YOLOX-S, as depicted in Figure 5.

2.4 Loss function

The loss function in the YOLOX-S model is derived by combining the bounding box loss, object classification loss, and confidence loss. The IoU loss function is the earliest one used in the bounding box loss function. It is computed by taking the intersection over union ratio of target box and anchor box. In order to accurately represent the relative positions of the two boxes, the GIoU loss function was proposed. This function considers the non-box area by setting the minimum bounding rectangle that encloses both target box and anchor box. It comprehensively considers the overlapping areas between target box and anchor box. However, if the anchor box completely contains the target box, the GIoU will be



equivalent to the IoU. So, the paper focuses on optimizing the accuracy of the loss function by introducing the EIoU bounding box loss function, which provides a comprehensive description of the positional relationship between target box and anchor box (Zhang, et al., 2022). Figure 6 illustrates the principle of EIoU calculation. The EIoU loss function comprises three components: overlap loss, center distance loss, and width and height loss between the target box and the anchor box, which are computed by Eq. 10.

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} \\ = 1 - IoU + \frac{\rho^2(b, b^{gt})}{(c_w)^2 + (c_h)^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \quad (10)$$

Where, b and b^{gt} represent the centroids of the anchor box and the target box, respectively; $\rho(\cdot)$ denotes the Euclidean distance between the two centroids; c_w and c_h refer to the width and height of the smallest outer rectangle of the target box and the anchor box, respectively; w and h represent the width and height of the anchor box, while w^{gt} and h^{gt} represent the width and height of the target box.

3 Experiment

3.1 Data set

In order to validate the effectiveness of the proposed model for recognizing the external force objects and smoke objects, transmission towers, excavators, bulldozers, concrete mixer, tower cranes, cranes, and engineering trucks; and smoke targets are selected as the recognition objects. Images captured by monitoring equipment in a specific province serve as the data source for constructing a sample library. This library comprises 15,485 images depicting the external force scenario of the transmission line corridor. We selected clear targets with diverse angles and backgrounds for labeling in the collected images. For smoke, smoke generated by a fire source was selected as a labeled target. A total of 4,003 images containing recognized object were labeled through object setting and screening, resulting in a total of

TABLE 2 A detailed description of image database of external force object.

Object	Label name	Number
Pole and Tower	ganta	2,587
Excavator	wajueji	2,183
Bulldozer	tuituji	849
Concrete mixer	jiaobanche	683
Tower crane	tadio	1,035
Crane	diaoche	840
Trucks	fandouche	1,359
Smoke	yanwu	593

10,129 labeled targets. The label names and corresponding numbers of these labeled targets are displayed in the Table 2.

The training and validation sets were derived from the image sample library, with 80% of the job images randomly assigned to the training set and the remaining 10% allocated to the validation set. Consequently, 3,202 images were used for training, while the validation set comprised 401 images. The rest of the images serve as a test set. In order to speed up training and prevent the weights from being corrupted, the model is trained with pre-trained weights and the 50 epochs are set to freeze training. The model was trained using python 3.7 and PyTorch on Ubuntu 18.04. The training was run on a single GeForce GTX 1080 with CUDA 10.1.

3.2 Visualization of attention maps

In order to demonstrate the enhanced effect on smoke which have random features, we utilize grad class activation map (Grad-CAM) for attention visualization (Selvaraju et al., 2017). The Grad-CAM is employed to depict the attention of target positions within

the convolutional layers. By pooling the average of gradients across the entirety of the final convolution layer, weights can be calculated for each channel. These weights are then applied to the feature map to generate a class activation map. The CAM assigns importance to each pixel in relation to the classification result. The Figure 7 depicts a comparison between the pre-improvement and post-improvement states (the darker middle portion of the figure indicates increased attention on the object). It is evident from the figure that the improved model extends attention to a wider range of features within the smoke region. This indicates that the improvement effectively enhances the model's feature extraction capability and enables improved extraction of irregular features.

3.3 Training result

The training process is shown in Figure 8. The loss function of the proposed model continuously decreases with iterative training, exhibiting a rapid decrease at the beginning and at the 50th epoch, then stabilizes around the 200th epoch, indicating good convergence performance. During the model training, the overall accuracy curve follows a similar trend as the loss function curve. The overall precision stabilizes around 200 epochs, reaching approximately 81.7%.

As shown in Figure 9, the experimental results demonstrate that the improved YOLOX-S achieves external force target recognition in complex scenarios. To validate the effectiveness of the improved model, a sample of images from the dataset is randomly selected and tested for detecting external force targets. The recognition results obtained using the original network and the improved network are presented below.

As depicted in the Figure 10, both YOLOX-S and the improved YOLOX-S successfully recognize the external force targets in scenario A. However, when compared to the YOLOX-S, the improved YOLOX-S exhibits superior confidence in detecting each target. Nonetheless, the results from the YOLOX-S network contain instances of missed detections. The missed targets

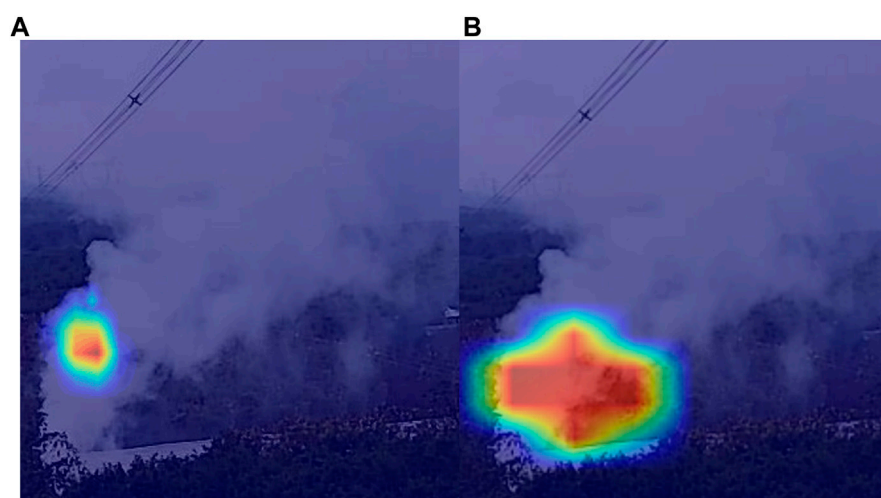


FIGURE 7
Grad-CAM attention map. (A) YOLOX-S, (B) Improved YOLOX-S.

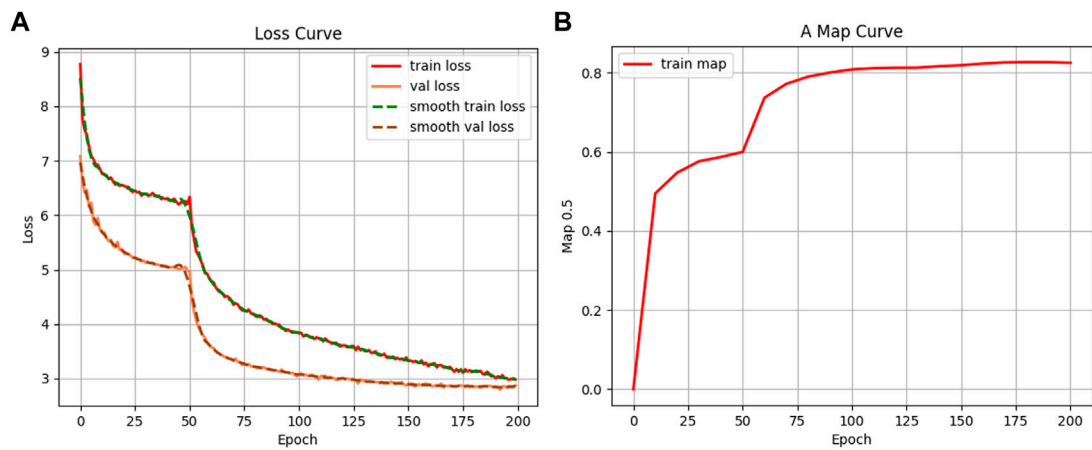


FIGURE 8
Variation curve of loss function (A) and accuracy curve (B) in training process.

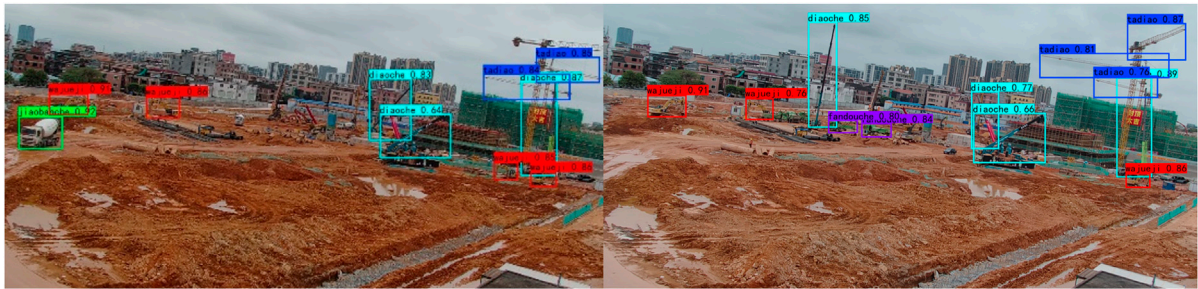


FIGURE 9
External force object detection result.

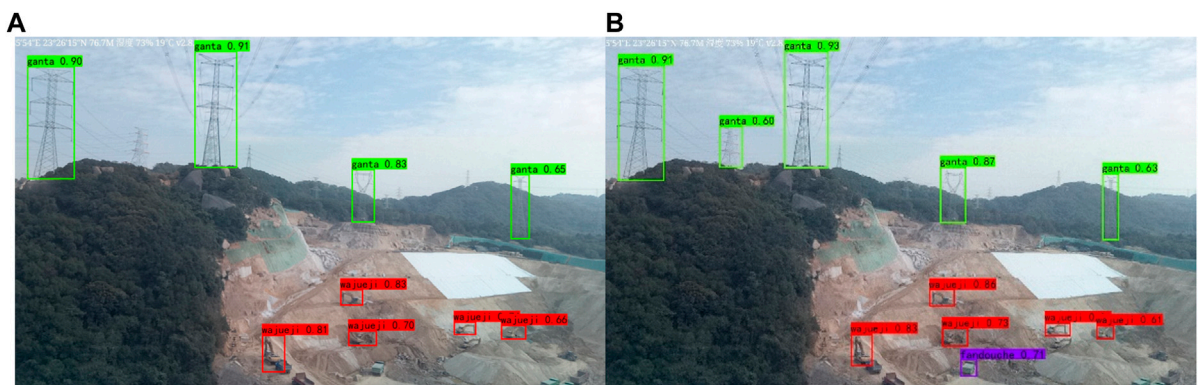
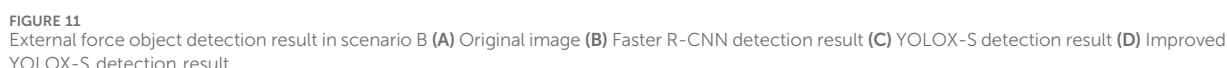


FIGURE 10
External force object detection result in scenario (A) YOLOX-S detection result (B) Improved YOLOX-S detection result.

correspond to smaller external force objects (trucks) or objects with color features resembling the environment (towers), resulting in missed detections. These findings highlight the efficacy of the global context block and attention module in enhancing target recognition.

For comparison, the two-stage network Faster R-CNN is selected (Ren, et al., 2015). In external force scenario B, both Faster R-CNN and the improved YOLOX-S successfully recognize all the targets in the Figure 11. However, the original network exhibits a missed detection for the smaller



To further evaluate the model’s recognition precision of targets in the transmission corridors, an ablation study is conducted. The ablation study compares the impact of each improvement component on object recognition precision. The Table 4 illustrates that GC block enhances network precision by 1.38%, while CBAM improves it by 2.24%. GC block enables targets in complex outdoor contexts to be attended to globally, slightly improving recognition accuracy. The attention mechanism improves the model recognition accuracy more, making the model pay more attention to the external broken target in both channel and space. Conversely, EIou has a minor impact on precision, affecting it by merely 0.45%. The EIou is mainly designed to describe the position of the anchor and target boxes well and provide more help to the model training process.

TABLE 3 Comparison of detection results under different target detection models.

Model	Average recognition precision of different external force targets (AP%)								mAP %
	Pole and tower	Excavator	Bulldozer	Concrete mixer	Tower crane	Crane	Trucks	Smoke	
Faster RCNN	80.65	87.37	88.21	77.36	83.45	76.26	76.37	58.13	78.47
YOLOX-S	81.34	85.16	87.11	76.73	84.63	75.46	74.78	57.72	77.87
YOLOv7	82.12	87.63	88.42	79.24	86.24	77.48	78.19	58.21	79.69
Improved YOLOX-S	83.43	90.77	91.96	78.44	88.93	79.35	79.91	61.11	81.74

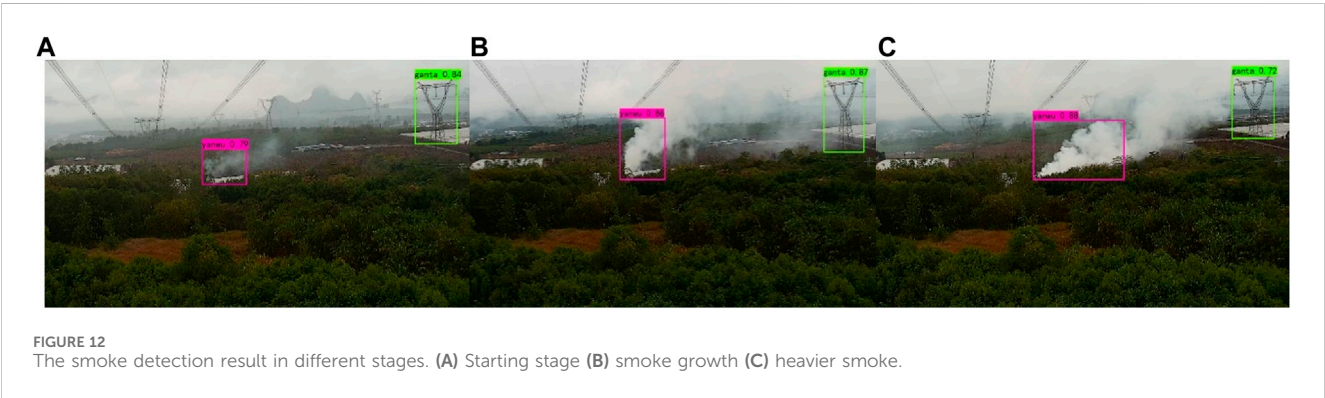


TABLE 4 Ablation study.

GC block	CBAM	EIoU	mAP%
\	\	\	77.87
✓	\	\	79.25
\	✓	\	80.11
✓	✓	\	81.29
✓	✓	✓	81.74

4 Conclusion

This study focuses on the recognition of external force targets in transmission line corridors, characterized by complex backgrounds, various object types, and irregular features. To achieve this, the model was improved. Conclusions can be drawn based on the model training and recognition results obtained from the dataset:

- (1) The improved YOLOX-S effectively identifies external force targets, exhibiting superior performance in complex environments and multi-target scenarios. The enhancements provided by the module improve global perception and target recognition capabilities, particularly for targets with distinct features. The recognition precision of such targets is enhanced by approximately 3% compared to the original model.

- (2) The improved network significantly enhances the recognition precision of smoke with random features. The attention heat map generated by Grad CAM demonstrates that the improved module effectively focuses on irregular targets, further refining the network's ability to recognize smoke with non-uniform characteristics. The model demonstrates improved performance in recognizing smoke at various stages and exhibits enhanced tracking capabilities for dynamically changing smoke.

This study incorporates the global context block, attention mechanism, and a new loss function to enhance the YOLOX-S network, resulting in an improved ability to recognize external force object. In light of the aforementioned research, the ensuing research will concentrate on two primary objectives. Firstly, to enhance the recognition precision of external force object and broaden the spectrum of recognizable types. Secondly, to develop an effective assessment of threats to external force transmission lines corridor and conduct an in-depth analysis of their operational environment. Specifically, it was found that the accuracy of smoke recognition needs to be further improved. With frequent changes in ambient wind direction, it is difficult to identify and localize the location of smoke generation. The situation is similar to natural fog, so there is a need to more clearly distinguish between fog and smoke and to more precisely locate smoke. We will analyze the effect on the light reflection effect in terms of the difference between the material composition and composition ratio of smoke and fog, and to use it as a new feature input to distinguish between smoke and fog. Next, we will use the number of various types of external force in the image, the distance and the working range of engineering machinery as the main transmission line external force threat assessment basis.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

HZ: Writing–review and editing. ZY: Writing–original draft. JS: Writing–review and editing. JC: Writing–review and editing. QY: Writing–review and editing. YC: Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research is supported by a project from Yunnan Provincial Department of

Science and Technology Major Science and Technology Special Projects (202202AD080004).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Cao, Y., Xu, J., Lin, S., Wei, F., and Hu, H. (2019). GCNet: non-local networks meet squeeze-excitation networks and beyond. *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), 11–17 Oct. 2021. IEEE, 1971–1980. doi:10.1109/ICCVW.2019.00246
- Cui, Y., Fang, C., Wen, Z., Fang, M., You, H., and Guo, J. (2023). Vibration signal identification of external force failure based on time-frequency spectrum and adaptive dynamic weight PSO-CNN algorithm. *Foreign Electron. Meas. Technol.* 42 (01), 144–152. doi:10.19652/j.cnki.femt.2204368
- Dong, Z., Gao, Y., Yuan, B., Yao, X., Zhang, J., and Zeng, J. (2023). An external damage detection method of transmission lines based on improved YOLOv4. *Power Syst. Clean Energy* 39 (06), 17–25. doi:10.3969/j.issn.1674-3814.2023.06.003
- Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). YOLOX: exceeding yolo series in 2021. *arXiv preprint*, 2021: arXiv:2107.08430. doi:10.48550/arXiv.2107.08430
- Golightly, I., and Jones, D. (2003). Corner detection and matching for visual tracking during power line inspection. *Image & Vis. Comput.* 21 (9), 827–840. doi:10.1016/S0262-8856(03)00097-0
- Gu, K., Xia, Z., Qiao, J., and Lin, W. (2020). Deep dual-channel neural network for image-based smoke detection. *IEEE Trans. Multimedia* 22 (2), 311–323. doi:10.1109/TMM.2019.2929009
- Hu, J., Xu, X., Ma, H., and Yan, Z. (2022). Distributionally robust Co-optimization of transmission network expansion planning and penetration level of renewable generation. *J. Mod. Power Syst. Clean Energy* 10 (3), 577–587. doi:10.35833/MPCE.2021.000156
- Krizhevsky, A., Ilya, S., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi:10.1145/3065386
- Larrauri, I. J., Sorrosal, G., and González, M. (2013). “Automatic system for overhead power line inspection using an Unmanned Aerial Vehicle-RELIFO project,” in *International Conference on Unmanned Aircraft Systems*, Atlanta, GA, June 13–14 2013 (IEEE), 244–252. doi:10.1109/ICUAS.2013.6564696
- Liang, Z. (2014). Statistical analysis of transmission line fault tripping in state grid corporation of China in 2011–2013. *East China Power* 42 (11), 2265–2270.
- Lin, G., Wang, B., Peng, H., Wang, X., Chen, S., and Zhang, L. (2019). Multi-target detection and location of transmission line inspection image based on improved Faster-RCNN. *Electr. Power Autom. Equip.* 39 (05), 213–218. doi:10.16081/j.issn.1006-6047.2019.05.032
- Liu, J., Huang, H., Zhang, Y., Luo, J., and He, J. (2019). Deep learning based external-force-damage detection for power transmission line. *2018 3rd Int. Conf. Commun. Image Signal Process. Sanya, China IOPscience* 1169 (1), 012032. doi:10.1088/1742-6596/1169/1/012032
- Long, L., Zhou, L., Liu, S., Huang, B., and Fan, K. (2022). Identification of hidden damage targets by external forces based on domain adaptation and attention mechanism. *J. Electronic Meas. Instrum.* 36 (11), 245–253. doi:10.13382/j.jemi.B2205638
- Lu, J., Zhou, T., Wu, C., Li, B., Tan, Y., and Zhu, Y. (2016). Fault statistic and analysis of 220kV and above power transmission line in province-level power grid. *High. Volt. Eng.* 42 (01), 200–207. doi:10.13336/j.1003-6520.hve.2016.01.026
- Ma, F., Wang, B., Dong, X., Yao, Z., and Wang, H. (2022). Safety image interpretation of power industry: basic concepts and technical framework. *Proc. CSEE* 42 (02), 458–475. doi:10.13334/j.0258-8013.psee.210315
- Ma, F., Wang, X., and Zhou, Y. (2021). Receptive field vision edge intelligent recognition for ice thickness identification of transmission line. *Power Syst. Technol.* 45 (06), 2161–2169. doi:10.13335/j.1000-3673.pst.2020.1015
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). “You only look once: unified, real-time object detection,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, NV, USA, 17 Jun, 2024 (IEEE), 779–788. doi:10.1109/CVPR.2016.91
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. *Adv. neural Inf. Process. Syst.* 39, 1137–1149. doi:10.1109/TPAMI.2016.2577031
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). “Grad-CAM: visual explanations from deep networks via gradient-based localization,” in *IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 2–3, 2023 (IEEE), 618–626. doi:10.1109/ICCV.2017.74
- Sheng, D., Deng, J., and Xiang, J. (2021). Automatic smoke detection based on SLIC-DBSCAN enhanced convolutional neural network. *IEEE Access* 9, 63933–63942. doi:10.1109/ACCESS.2021.3075731
- Tian, E., Li, C., Zhu, G., Su, Z., Zhang, X., and Xu, X. (2021). Identification algorithm of transmission line external hidden danger based on YOLOv4. *Comput. Syst. Appl.* 30 (7), 190–196. doi:10.15888/j.cnki.csa.008082
- Tian, K., Sun, W., and Han, D. (2022). Strategic investment in transmission and energy storage in electricity markets. *J. Modern Power Syst. Clean Energy* 10 (1), 179–191. doi:10.35833/MPCE.2020.000927
- Wang, B., Ma, F., Dong, X., Wang, P., Ma, H., and Wang, H. (2019). Electric power depth vision: basic concepts, key technologies and application scenarios. *Guangdong Electr. Power* 32 (09), 3–10. doi:10.3969/j.issn.1007-290X.2019.009.001
- Wang, B., Ma, F., Ge, L., Ma, H., Wang, H., Mohamed, M. A., et al. (2021). Icing-EdgeNet: a pruning lightweight edge intelligent method of discriminative driving channel for ice thickness of transmission lines. *IEEE Trans. Instrum. Measurement* 70, 1–12. Art no. 2501412. doi:10.1109/TIM.2020.3018831
- Wang, C. Y., Bochkovskiy, A., and Liao, H. Y. M. (2023). “YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, Fri Jun 21st, 2024 (IEEE), 7464–7475. doi:10.1109/CVPR52729.2023.00721
- Wei, Y., Li, M., Xie, Y., and Dai, B. (2022). Transmission line inspection image detection based on improved Faster-RCNN. *Electr. Power Eng. Technol.* 41 (02), 171–178. doi:10.12158/j.2096-3203.2022.02.023

- Wei, Z., Lu, W., Zhao, W., and Wang, D. (2021). Target detection method for external damage of a transmission line based on an improved Mask R-CNN algorithm. *Power Syst. Prot. Control* 49 (23), 155–162. doi:10.19783/j.cnki.pspc.210482
- Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). “Cbam: convolutional block attention module,” in *Proceedings of the European conference on computer vision (ECCV)* (Munich, Germany: Springer), 3–19. doi:10.48550/arXiv.1807.06521
- Wu, M., Guo, L., Chen, R., Du, W., Wang, J., Liu, M., et al. (2022). Improved YOLOX foreign object detection algorithm for transmission lines. *Wirel. Commun. Mob. Comput.* 2022, 1–10. doi:10.1155/2022/5835693
- Xiao, X., and Zheng, Z. (2022). New power systems dominated by renewable energy towards the goal of emission peak & carbon neutrality: contribution, key techniques, and challenges. *Adv. Eng. Sci.* 54 (1), 46–59. doi:10.15961/j.jsuese.202100656
- Zhang, A., and Deng, F. (2020). External vibration identification scheme of transmission line tower based on CNN-rvm. *Comput. Simul.* 36 (04), 76–80.
- Zhang, J., Yu, J., Wang, J., and Tan, S. (2018). Image recognition technology for transmission line external damage based on depth learning. *Comput. Syst. Appl.* 27 (8), 176–179. doi:10.15888/j.cnki.csa.006458
- Zhang, Y., Ren, W., Zhang, Z., Jia, Z., Wang, L., and Tan, T. (2022). Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* 506, 146–157. doi:10.1016/j.neucom.2022.07.042



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Jasronita Jasni,
Putra Malaysia University, Malaysia
Huakun Huang,
Guangzhou University, China

*CORRESPONDENCE

Xiaojing Bai,
✉ baixiaojing@ncepu.edu.cn

RECEIVED 24 November 2023

ACCEPTED 06 February 2024

PUBLISHED 21 February 2024

CITATION

Zhang Q, Peng G, Tan Y, Zhang Z and Bai X (2024), A transformer acoustic signal analysis method based on matrix pencil and hybrid deep neural network.

Front. Energy Res. 12:1343748.

doi: 10.3389/fenrg.2024.1343748

COPYRIGHT

© 2024 Zhang, Peng, Tan, Zhang and Bai. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A transformer acoustic signal analysis method based on matrix pencil and hybrid deep neural network

Qizhe Zhang¹, Guozheng Peng¹, Yuanpeng Tan¹,
Zhonghao Zhang¹ and Xiaojing Bai^{2*}

¹Artificial Intelligence Application Research Department, China Electric Power Research Institute, Beijing, China, ²North China Electric Power University, Beijing, China

Acoustic signal analysis is an important component of transformer online monitoring. Currently, traditional methods have problems such as low spectral resolution, imbalanced sample distribution, and unsatisfactory classification performance. This article first introduces the matrix pencil algorithm for time-frequency spectrum analysis of acoustic signals, and then uses the SMOTE algorithm to expand the imbalanced samples. Then, an ACmix hybrid deep neural network model is constructed to classify 11 types of transformer operation and environmental acoustic signals. Finally, detailed experiments were conducted on the method proposed in this paper, and the experimental results showed that the matrix pencil algorithm has high time-frequency resolution and good noise resistance performance. The SMOTE sample expansion method can significantly improve the recognition accuracy by more than 2%. Overall accuracy of the proposed method in acoustic signal classification tasks reaches 91.81%.

KEYWORDS

matrix pencil, transformer acoustic signal, deep neural network, sample expansion, attention mechanism

1 Introduction

As a high-value key equipment in the power system, power transformers need to undergo maintenance work throughout their entire life cycle to ensure the healthy operation (Wardani et al., 2011; Liang et al., 2017). During the operation of transformers, mechanical waves are generated due to the vibration of the iron core and winding, which in turn generate sound waves through the fluid medium (Hsu et al., 2015; Wang et al., 2021). Furthermore, when the insulation of transformers deteriorate, partial discharge may occur (Okabe and Wada, 2011), which can also trigger acoustic signals. Acoustic signals near transformers contain a large amount of equipment status information (Cole, 1997). The collection of acoustic signals has advantages such as non-contact, low cost, and convenience, and thus has broad application prospects in the field of online monitoring of transformers (Sithole et al., 2019; Kucera et al., 2022).

In terms of feature extraction of acoustic signals, it mainly includes time-domain feature extraction, frequency-domain feature extraction, and time-frequency domain feature extraction (Zhang et al., 1998; Lee et al., 2021; Caldeira and Coelho, 2023). The time-frequency spectrum contains both time-domain and frequency-domain information, which has better performance in fault identification (Geng et al., 2019). In the field of pattern

recognition of acoustic signals, traditional machine learning methods such as support vector machine (SVM) (Wu et al., 2018), back propagation neural network (BP NN) (Wang et al., 2011) have gradually developed deep learning methods such as convolutional neural network (Lu et al., 2019; Kollias and Zafeiriou, 2021).

With the rapid development of artificial intelligence technology, edge computing technology (Huang et al., 2023), intelligent perception (Zhao et al., 2024) and other technologies, online monitoring of equipment status has been widely used. But the effectiveness of related applications still needs to be improved. The pattern recognition algorithm based on deep neural networks faces the problem of low accuracy in engineering applications. Reference (Huang et al., 2020) introduces Gaussian Bernoulli restricted Boltzmann machines to improve algorithm performance. In different application fields, data preprocessing methods and pattern recognition algorithms require specific improvements to adapt to corresponding tasks.

Currently, there are still three issues in the analysis of transformer acoustic signals:

- The spatiotemporal resolution of the time-frequency spectrum is insufficient, and there is a problem of signal interference.
- Treating environmental sounds such as bird singing and rain sound as background noise, ignoring the risk factors present in the transformer environment;
- The distribution of acoustic signal samples is uneven, and the recognition accuracy is unsatisfactory.

This article focuses on the above issues and introduces the matrix pencil algorithm to improve the time-frequency resolution of acoustic signals and enhance the noise resistance performance. Expand imbalanced samples using the SMOTE algorithm (Bao and Yang, 2023). An ACmix model was constructed to classify transformer operation and environmental sound. Finally, the effectiveness of the method proposed in this article was verified.

2 Research methodology

2.1 Transformer acoustic signal time-frequency spectrum

Spectrum is an important feature in the analysis and processing of acoustic signals. It can reflect the frequency energy distribution of signals at different times, establish a good connection between the time and frequency domains of acoustic signals, and achieve the maximization of acoustic feature information. The generation of time-frequency spectrum mainly includes the process of framing, windowing, and frequency analysis. Traditional frequency analysis methods of acoustic signals are mainly based on Fourier transform, which has problems such as low accuracy and poor anti-interference ability. In contrast, the matrix pencil method (Wang et al., 2017) belongs to the subspace rotation invariant method, which utilizes the orthogonal characteristics of the signal subspace to construct spectral peaks, thereby improving the frequency resolution and anti-interference performance of the algorithm. And since there

is no need for iterative operations, the efficiency is high. This article uses matrix pencil method to construct the time-frequency spectrum of sound signals.

2.1.1 Acoustic signal model

The transformer acoustic signal model can be represented as Eq. 1.

$$s(n) = \sum_{k=1}^K a_k e^{-\beta_k n \Delta t} \cos(n \omega_k \Delta t + \theta_k) \quad (1)$$

Among them, α_k , ω_k , and θ_k represent the amplitude, frequency, and initial phase of the k -th frequency component, respectively. β_k is the attenuation coefficient of the k -th frequency component. Δt is the sampling time interval. Further represented as an exponential model in Eq. 2.

$$s(n) = \sum_{k=1}^{2K} R_k e^{p_k n \Delta t} \quad (2)$$

Among them, $R_k = \alpha_k e^{\pm j\theta/2}$; $p_k = -\beta_k \pm j\omega_k$.

Considering the presence of noise $u(n)$, the actual observed signal is described in Eq. 3.

$$x(n) = s(n) + u(n) \quad (3)$$

2.1.2 Matrix pencil method

For the acoustic signal $x(n)$, construct two $(N-L) \times L$ Hankel matrices X_1 and X_2 are represented as Eqs 4, 5.

$$X_1 = \begin{bmatrix} x(0) & x(1) & \cdots & x(L-1) \\ x(1) & x(2) & \cdots & x(L) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-L-1) & x(N-L) & \cdots & x(N-2) \end{bmatrix} \quad (4)$$

$$X_2 = \begin{bmatrix} x(1) & x(2) & \cdots & x(L) \\ x(2) & x(3) & \cdots & x(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-L) & x(N-L+1) & \cdots & x(N-1) \end{bmatrix} \quad (5)$$

Where, L is the matrix pencil parameter, and appropriate selection can reduce the impact of noise. $N/3$ to $2N/3$ is the better choice for L . Represent X_1 and X_2 as Eqs 6, 7

$$X_1 = Z_L P Z_R \quad (6)$$

$$X_2 = Z_L P Z Z_R \quad (7)$$

The variables in Eqs 6, 7 are calculated based on Eqs 8–11.

$$Z_L = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{p_1 \Delta t} & e^{p_2 \Delta t} & \cdots & e^{p_{2K} \Delta t} \\ \vdots & \vdots & \ddots & \vdots \\ e^{p_1 (N-L-1) \Delta t} & e^{p_2 (N-L-1) \Delta t} & \cdots & e^{p_{2K} (N-L-1) \Delta t} \end{bmatrix} \quad (8)$$

$$Z_R = \begin{bmatrix} 1 & e^{p_1 \Delta t} & \cdots & e^{p_1 (L-1) \Delta t} \\ 1 & e^{p_2 \Delta t} & \cdots & e^{p_2 (L-1) \Delta t} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{p_{2K} \Delta t} & \cdots & e^{p_{2K} (L-1) \Delta t} \end{bmatrix} \quad (9)$$

$$Z = \text{diag}[e^{p_1 \Delta t}, e^{p_2 \Delta t}, \dots, e^{p_{2K} \Delta t}] \quad (10)$$

$$P = \text{diag}[R_1, R_2, \dots, R_{2K}] \quad (11)$$

Then the following Eq. 12 is obtained, i.e.,

$$X_1 - \lambda X_0 = Z_L P (Z_0 - \lambda I) Z_R \quad (12)$$

The generalized eigenvalue of matrix pencil $X_1 - \lambda X_0$ includes the number, frequency, and attenuation coefficient of frequency components. Thus the calculation of signal frequency components can be transformed into solving Eq. 13 for generalized features:

$$X_1 - \lambda X_0 = X_0^+ X_1 \quad (13)$$

In the equation, X^+ is the pseudo inverse of X_0 .

After obtaining the number, frequency, and attenuation coefficient of frequency components of acoustic signals, the signal amplitude can be obtained by solving the following least squares problem shown in Eq. 14.

$$\begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{p_1 \Delta t} & e^{p_2 \Delta t} & \cdots & e^{p_{2K} \Delta t} \\ \vdots & \vdots & \ddots & \vdots \\ e^{p_1 (N-1) \Delta t} & e^{p_2 (N-1) \Delta t} & \cdots & e^{p_{2K} (N-1) \Delta t} \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_{N-1} \end{bmatrix} \quad (14)$$

2.2 Imbalanced sample expansion method

The acoustic signal samples of transformers in different states exhibit significant imbalance, especially in the absence of abnormal samples. To address the issue of sample imbalance, SMOTE is used for sample expansion.

The SMOTE algorithm is an algorithm for expanding minority class samples in imbalanced samples, which is an improvement on the random oversampling algorithm. The random oversampling algorithm expands the dataset by simply copying minority samples, which is prone to overfitting and other issues. The SMOTE algorithm combines the ideas of nearest neighbor analysis and linear interpolation.

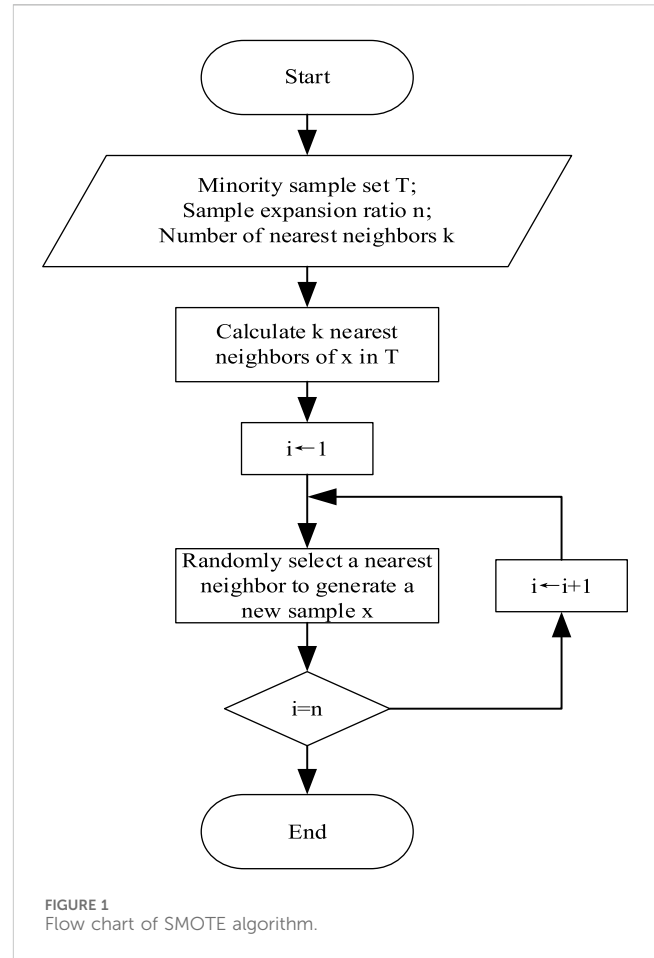
Firstly, traverse the minority class samples, select each minority class sample A as the reference value for a single expansion, and calculate its Euclidean distance from other sample point B in the minority class samples according to Eq. 15

$$d(A, B) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (a_{i,j} - b_{i,j})^2} \quad (15)$$

Obtain k nearest neighbors of sample point A based on the size of Euclidean distance, randomly select a certain nearest neighbor B, and perform linear interpolation with reference sample A according to Eq. 16 to generate a new sample C. Repeat this n times (n is the given sample expansion ratio)

$$c_{i,j} = a_{i,j} + (b_{i,j} - a_{i,j}) \cdot \theta \quad (16)$$

In Eq. 16, θ is a random number between (0, 1). After SMOTE process, for each sample A, n new samples will be generated. After traversing all the minority samples, the original number of samples will be expanded to n times the original number, achieving sample expansion through oversampling. The specific flowchart of SMOTE algorithm is shown in Figure 1.



2.3 Hybrid deep neural network

Convolutional neural network is composed of convolutional layer, pooling layer, and fully connected layer. Compared with traditional artificial neural networks, convolutional neural network has fewer connections between neurons in different layers. Convolutional layers learn features from network inputs, and different convolutional kernels have different feature extraction effect.

Attention mechanism is a special plug and play structure often embedded in convolutional neural networks, used to automatically learn and calculate the contribution of input data to output data. The self attention mechanism is a variant of the attention mechanism, which reduces dependence on external information and is better at capturing internal correlations of data or features. Attention mechanisms are also widely used in visual tasks. Compared to traditional convolutional models, attention allows the model to focus on important regions over a larger range.

This article combines the advantages of residual neural network and self attention mechanism, and uses ACmix algorithm to construct a transformer operation acoustic signal analysis model.

The schematic diagram of ACmix is shown in Figure 2:

First, three 1×1 convolutions are used to project the input features, and then reshape them into N pieces, thereby obtaining a set of $3 \times N$ intermediate features.

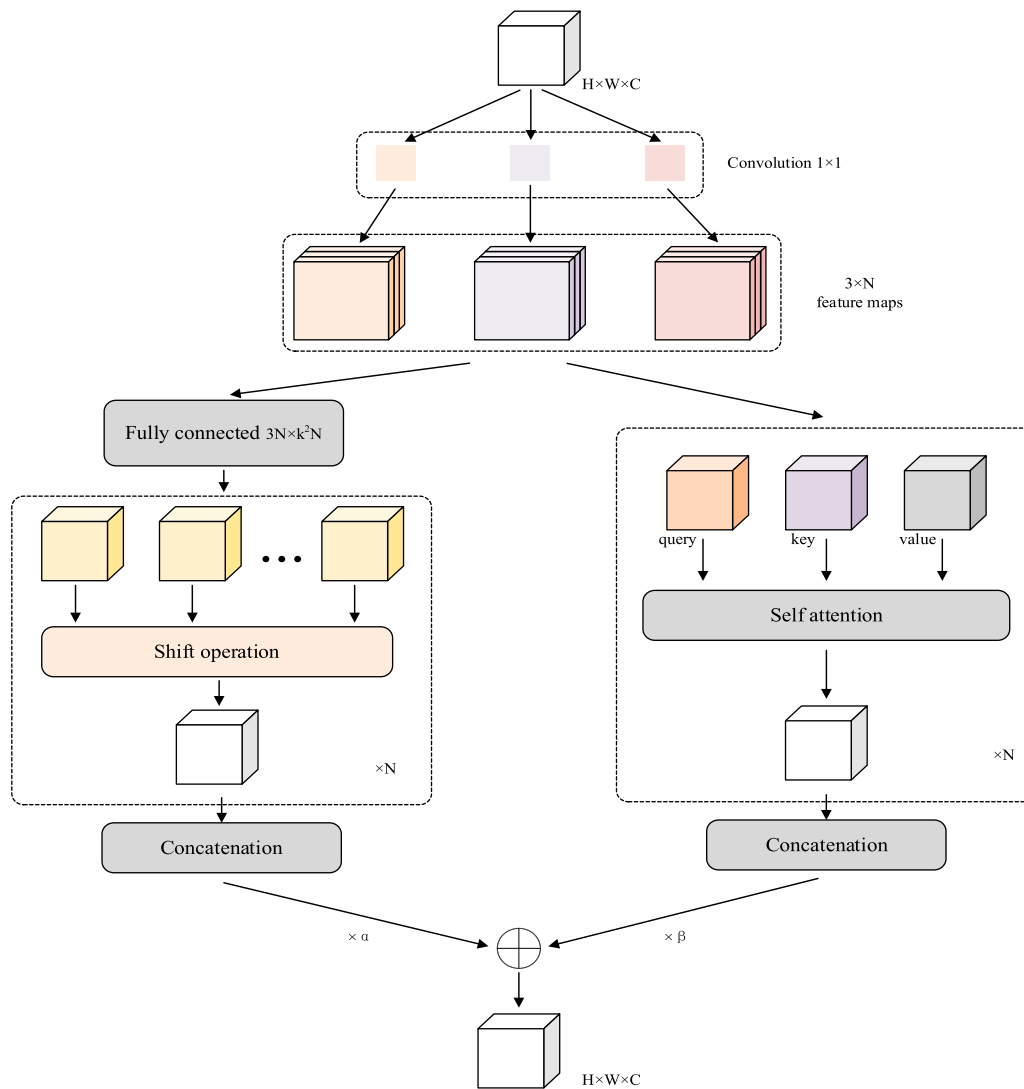


FIGURE 2
Schematic diagram of ACmix.

Secondly, parallel self attention computation and convolution computation are carried out. For the self attention part, gather the intermediate features into N groups, each containing 3 features, each from 1×1 convolution. The corresponding three feature maps are used as query, key, and value, respectively, using a multi head self attention module. The calculation formula is shown in Eqs 17, 18

$$q_{i,j}^{(l)} = W_q^{(l)} f_{i,j}, k_{i,j}^{(l)} = W_k^{(l)} f_{i,j}, v_{i,j}^{(l)} = W_v^{(l)} f_{i,j} \quad (17)$$

$$g_{i,j} = \parallel_{l=1}^N \left(\sum_{a,b \in N_{k(i,j)}} A(q_{i,j}^{(l)}, k_{a,b}^{(l)}) v_{a,b}^{(l)} \right) \quad (18)$$

where \parallel is the concatenation of the outputs of N attention heads. q , k , v represent the projection matrices for queries, keys and values. N represents the local area near the pixel point (i,j) with a range of k . A is the attention weight, f is the input, and g is the output.

In the convolutional part, a convolutional kernel with size k is used to generate k^2 feature maps using a light fully connected layer.

By shifting and aggregating the generated features, convolution processing is performed on the input features, and information is collected from the local receptive field. Finally, the output is obtained by Eq. 19

$$F_{out} = \alpha F_{att} + \beta F_{conv} \quad (19)$$

Based on the structure of ResNet50, replace the convolutional layer with ACmix module.

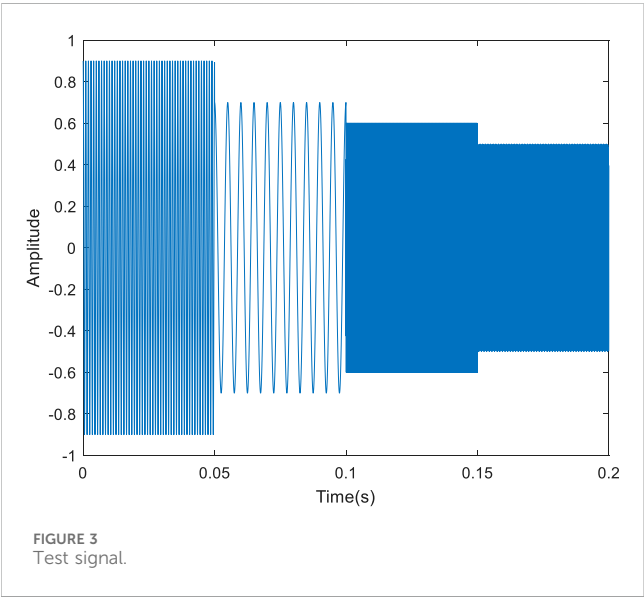
3 Manuscript experimental results and analysis

3.1 Performance verification of matrix pencil method

To verify the time-frequency spectrum analysis performance of matrix pencil method. Construct the test signal as shown in Table 1,

TABLE 1 Test signal parameters.

No.	Frequency/ kHz	Amplitude/ A	Phase	Duration/ s
1	1.00	0.90	0	0–0.05
2	0.20	0.70	0	0.05–0.10
3	18.00	0.60	0	0.10–0.15
4	5.00	0.50	0	0.15–0.20



and perform time-frequency spectrum analysis using matrix pencil method and Fourier transform, respectively.

The signal sampling rate is set to 48 kHz, and the waveform of the test signal is shown in Figure 3.

Set the time window to 0.01 s, and the moving step of the time window is 0.01 s. The time-frequency spectrum of the test signal was calculated using matrix pencil method and Fourier transform, respectively, as shown in Figure 4.

It can be seen that the frequency resolution of the Fourier method is relatively low. When the signal frequency changes, there are significant unclear areas in the time-frequency spectrum. The matrix pencil method can more accurately extract the frequency characteristics of signals, especially to accurately identify changes in signal frequency. To further verify the noise immunity of the matrix pencil method, Gaussian colored noise is added to the original signal, and the signal-to-noise ratio is set to 10. The test signal containing noise is shown in Figure 5.

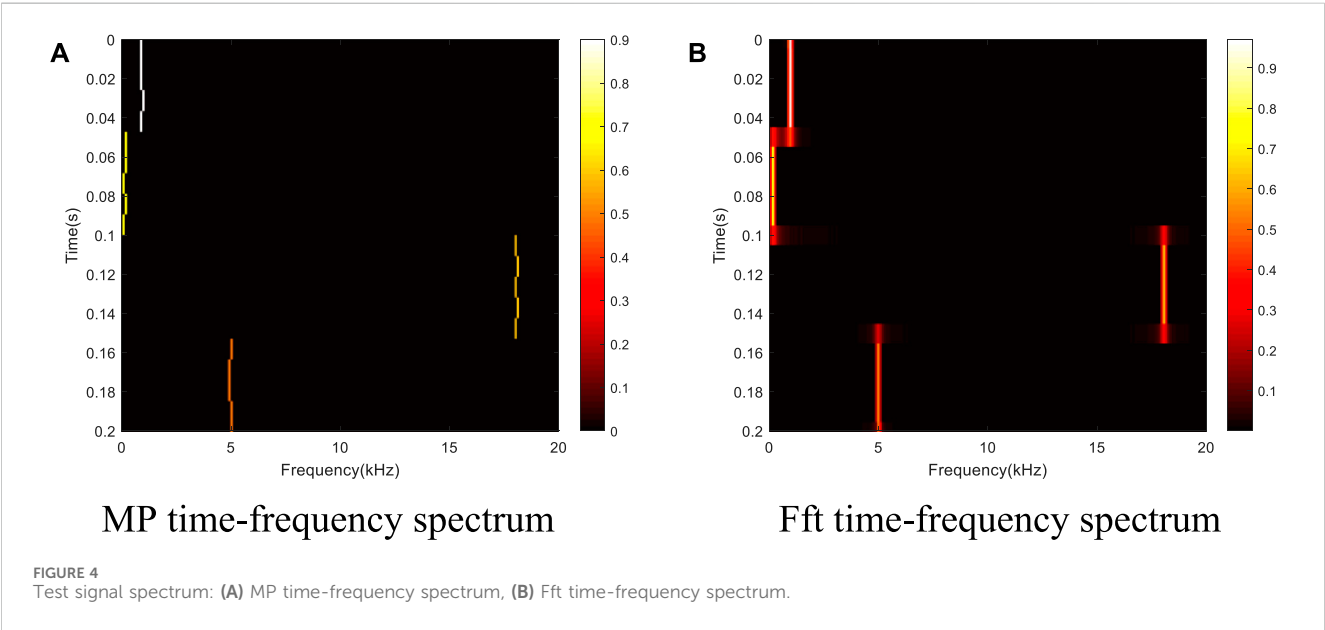
The time window and step size remain unchanged, and the matrix pencil method and Fourier transform are used to calculate the time-frequency spectrum of the test signal, as shown in Figure 6.

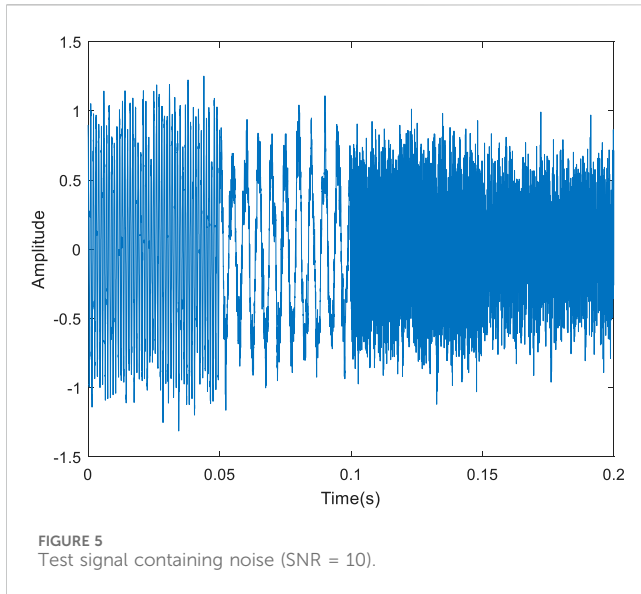
When noise exists, significant noise appears in the time-frequency spectrum obtained by the Fourier method. In contrast, the time-frequency spectrum background obtained by matrix pencil method is pure. Therefore, the matrix pencil method has higher resolution and stronger anti-interference ability compared to the Fourier method in the time-frequency spectrum analysis of acoustic signals.

3.2 Dataset and evaluation indicators

3.2.1 Sample information

2236 audio signals of various types were collected at substations of different voltage levels. The sampling rate is 48 kHz. Each audio segment has a collection time of 5 s. The dataset used in this article includes 11 types of acoustic signals, including normal operation of transformers with 5 voltage levels, bird singing, rain, OLTC switching, short circuit impulse, partial discharge, DC bias, etc. Encode it and use the SMOTE method for sample expansion. With the goal of balancing the number of expanded samples, expand the





original sample size to an integer multiple. After sample expansion, 4,524 samples were obtained. The sample information is shown in Table 2.

3.2.2 Acoustic signal preprocessing

Due to the different durations of characteristic signals in various states, it is necessary to intercept audio signals. After analysis, 2 s was selected as the duration for signal analysis. Intercept typical data with a duration of 2 s from the audio and convert the acoustic signal into a time-frequency spectrum using the method described in Chapter 2.1. The sampling rate of the sample data is 48 kHz. The width of the sliding window is 0.01 s, and the step length of the time window movement is 0.01 s. The frequency range of the time-frequency spectrum is 0–20 kHz, and the frequency resolution is set to 0.1 kHz. Convert the amplitude to a decibel value according to Eq. 20.

$$s_{db} = 20 \log_{10} s \quad (20)$$

Where s represents amplitude. The typical acoustic signal time-frequency spectrum of transformer operation is shown in Figure 7.

Subsequently, the time-frequency spectrum of the acoustic signal will be used as input data for classification processing using the deep neural network model.

3.2.3 Evaluation indicators

Evaluate the recognition rate of the model using accuracy, recall, F_1 score, and overall accuracy. The accuracy calculation formula is shown in Eq. 21.

$$p = \frac{x_{pt}}{x_p} \times 100\% \quad (21)$$

Among them, x_{pt} is the number of samples predicted to be x and actually x . X_p is the number of samples predicted to be x . The recall rate is calculated according to the formula Eq. 22.

$$r = \frac{x_{pt}}{x_t} \times 100\% \quad (22)$$

X_t is the actual number of samples that are x . The F_1 score is calculated by Eq. 23.

$$S_{F1} = 2 \frac{p \times r}{p + r} \times 100\% \quad (23)$$

The overall accuracy is defined as the proportion of the total number of correctly predicted samples to the total number of samples.

3.3 Comparison of classification effects

To verify the effectiveness of the SMOTE sample expansion method, the ACmix model was used to train and test the original and expanded samples, respectively. Epoch number was set to 300, and learning rate was set to 0.01. When using the original samples, the

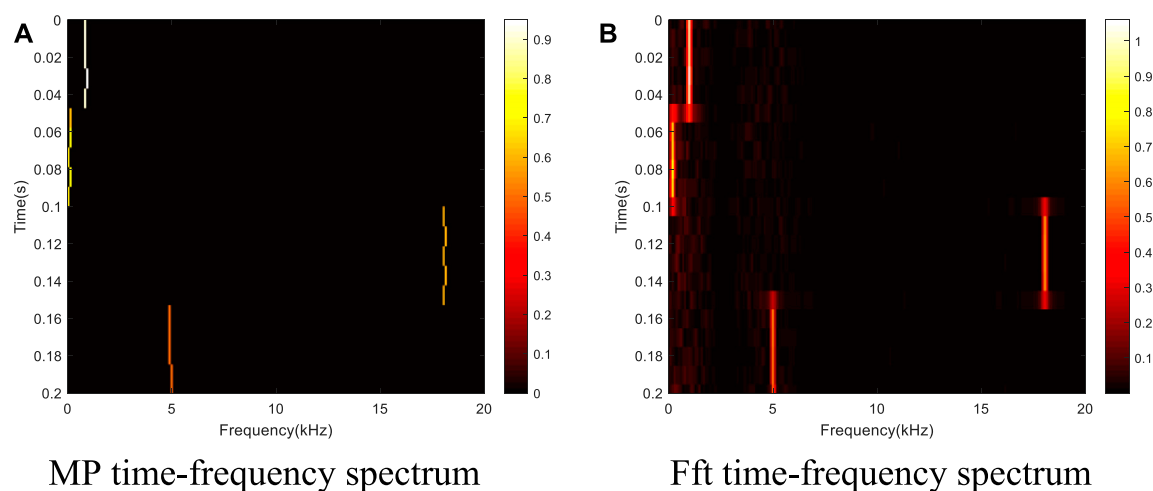
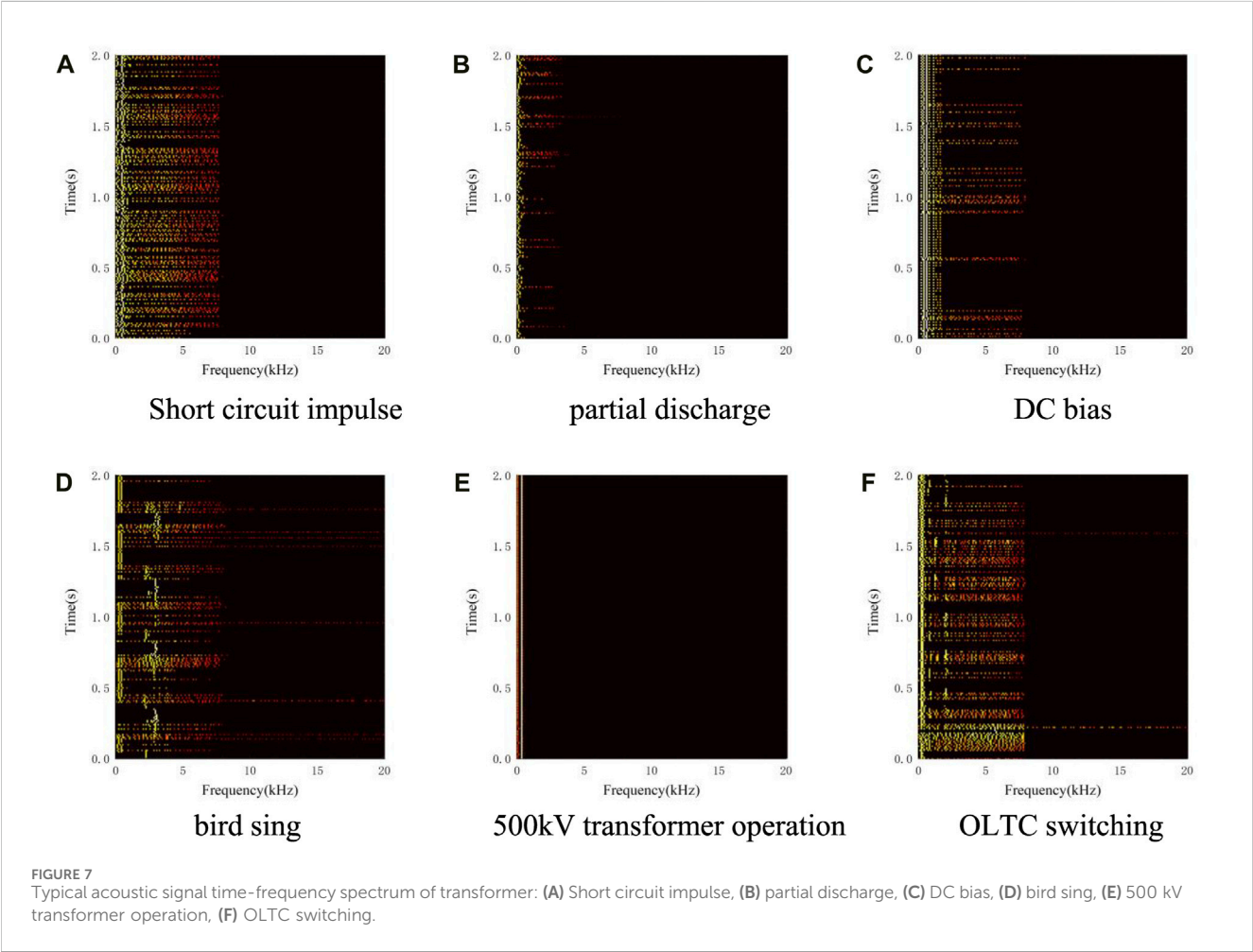


FIGURE 6
Time-frequency spectrum of test signal containing noise: (A) MP time-frequency spectrum, (B) Fft time-frequency spectrum.

TABLE 2 Sample information.

Code	Type	Original sample size	Expanded sample size
0	110 kV transformer operation	200	400
1	220 kV transformer operation	400	400
2	500 kV transformer operation	400	400
3	800 kV transformer operation	400	400
4	1000 kV transformer operation	450	450
5	Bird sing	60	420
6	Rain	48	384
7	OLTC switching	56	392
8	Short circuit impulse	42	378
9	Partial discharge	90	450
10	DC bias	90	450

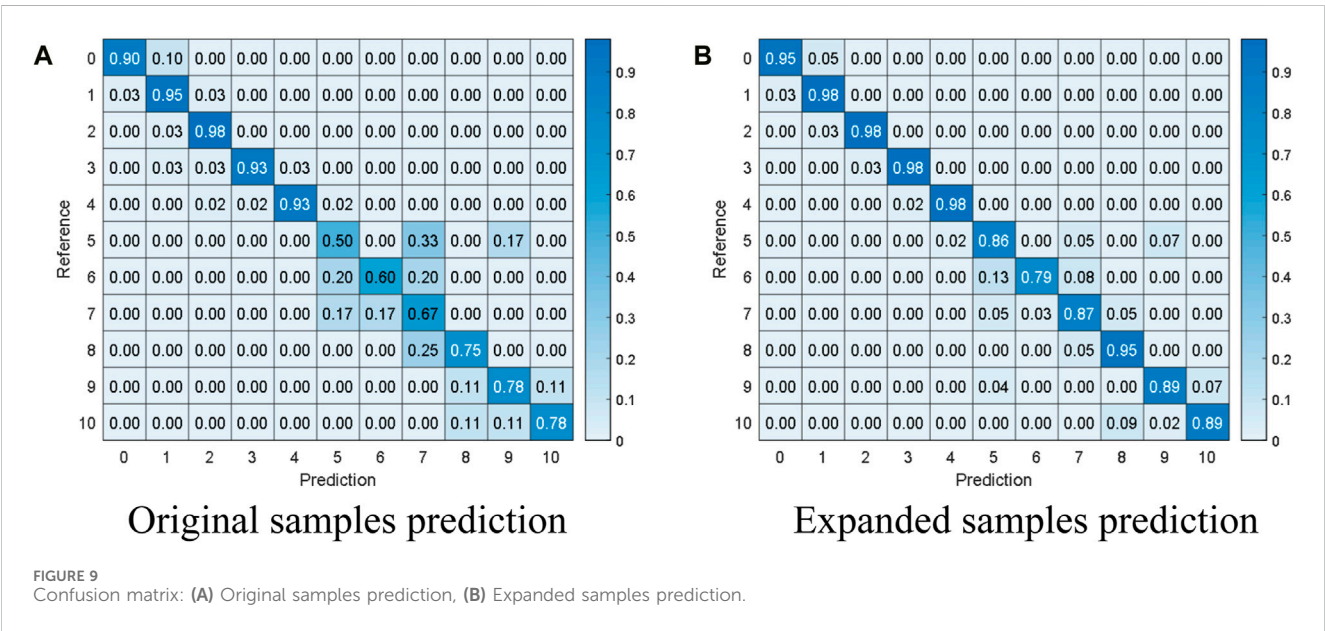
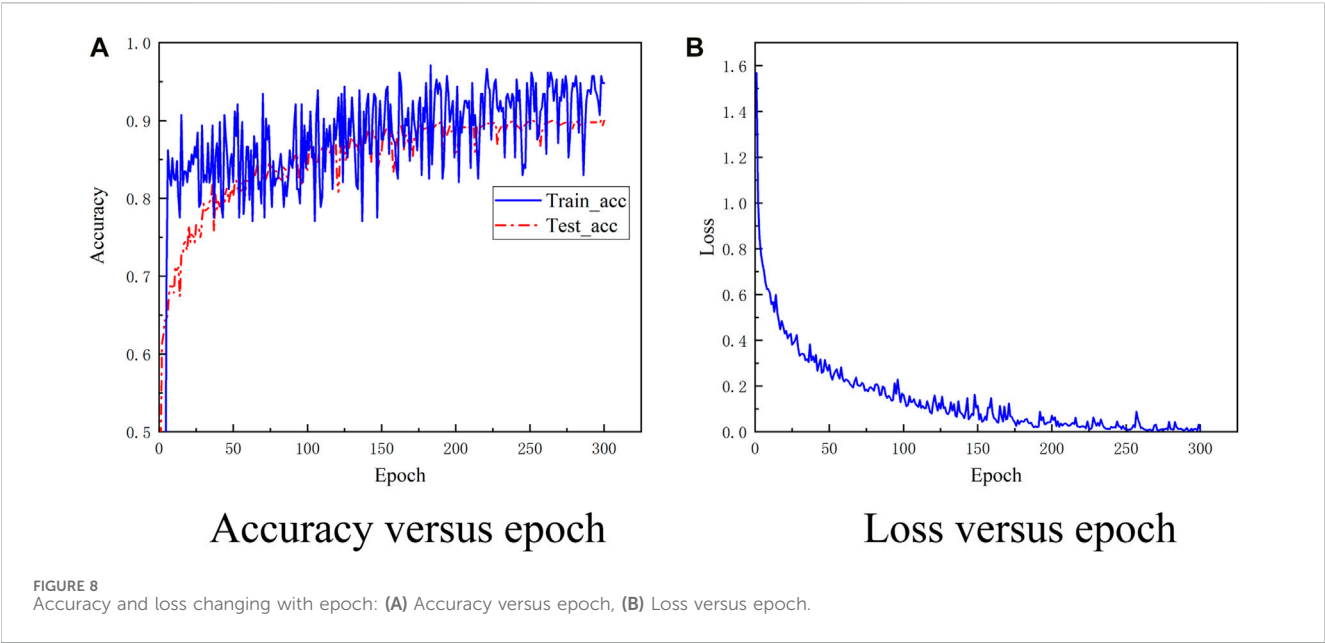
Divide the original sample and the expanded sample into training, testing, and validation sets in a ratio of 8:1:1.



accuracy and loss changes during the training process of ACmix model are shown in Figure 8.

The loss converges at 200th epoch. Therefore, in the subsequent training, the epoch number was set to 200. The confusion matrix of the validation set under two sample sets is shown in Figure 9.

After expanding the sample, the recognition accuracy of the sample was significantly improved. Especially, the accuracy of



minority class such as 6, 7, 8, and 9 have been significantly improved. Therefore, the SMOTE algorithm can effectively improve the classification performance of transformer samples.

Compare the ResNet50 model with ACmix and compare the performance of the model with the addition of the SMOTE algorithm. The classification results for the four scenarios are shown in Table 3.

For the ResNet50 model, the original accuracy was 83.48%, and after SMOTE sample expansion, the accuracy improved by 2.36%. The F_1 scores of each category have significantly improved, especially for samples in categories 6, 7, and 8,

with an increase of over 20%. For the ACmix model, the original accuracy was 89.73%, and after SMOTE sample expansion, the accuracy was improved by 2.08%. In categories 6, 7, 8, and 9, the F_1 score has increased by 20%. Due to the uneven distribution of the original samples, there is not much difference in overall accuracy before and after using the SMOTE algorithm to expand the samples. However, after using the SMOTE algorithm, the accuracy of the minority categories is significantly improved. The ACmix model has higher recognition accuracy for transformer operation acoustic signals compared to the ResNet50 model.

TABLE 3 Classification results in different scenarios.

Type	ResNet50			ResNet50 + SMOTE			ACmix			ACmix + SMOTE		
	p	r	S_{F1}	p	r	S_{F1}	p	r	S_{F1}	p	r	S_{F1}
0	0.85	0.85	0.85	0.97	0.95	0.96	0.95	0.90	0.92	0.97	0.95	0.96
1	0.85	0.88	0.86	0.93	0.98	0.95	0.90	0.95	0.93	0.93	0.98	0.95
2	0.84	0.90	0.87	0.98	0.98	0.98	0.93	0.98	0.95	0.98	0.98	0.98
3	0.92	0.88	0.90	0.98	0.98	0.98	0.97	0.93	0.95	0.98	0.98	0.98
4	0.95	0.87	0.91	0.98	0.96	0.97	0.98	0.93	0.95	0.98	0.98	0.98
5	0.43	0.50	0.46	0.67	0.69	0.68	0.50	0.50	0.50	0.80	0.86	0.83
6	0.50	0.40	0.44	0.96	0.71	0.82	0.75	0.60	0.67	0.97	0.79	0.87
7	0.38	0.50	0.43	0.67	0.72	0.69	0.50	0.67	0.57	0.83	0.87	0.85
8	0.60	0.75	0.67	0.76	0.89	0.82	0.60	0.75	0.67	0.86	0.95	0.90
9	0.88	0.78	0.82	0.79	0.82	0.80	0.78	0.78	0.78	0.91	0.89	0.90
10	0.78	0.78	0.78	0.83	0.78	0.80	0.88	0.78	0.82	0.93	0.89	0.91
Overall accuracy	83.48			85.84			89.73			91.81		

4 Conclusion

This article proposes a method for analyzing transformer acoustic signals. Firstly, the matrix pencil algorithm is introduced for time-frequency spectrum analysis of acoustic signals. The experimental results show that the matrix pencil algorithm has a significant advantage in time-frequency resolution and has good noise resistance performance. Secondly, to address the issue of imbalanced sample distribution, the SMOTE algorithm is used to expand the minority sample categories. Finally, an ACmix model was constructed to classify the operation and environmental sound of 11 types of transformers. The results show that SMOTE sample expansion can significantly improve the recognition accuracy of minority sample categories, and the ACmix model has good recognition performance. Research content of this article will provide technical support for intelligent analysis of transformer sound signals.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

QZ: Methodology, Writing–original draft. GP: Data curation, Writing–review and editing. YT: Supervision, Writing–review and

editing. ZZ: Validation, Writing–review and editing. XB: Supervision, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The work is funded by the science and technology project of SGCC (State Grid Corporation of China): Intelligent Management and Control of Power Material Inspection, Storage, Distribution Based on Multi-Perception and Edge-computing (5700-202255448A-2-0-ZN).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bao, Y., and Yang, S. (2023). Two novel SMOTE methods for solving imbalanced classification problems. *IEEE Access* 11, 5816–5823. doi:10.1109/ACCESS.2023.3236794
- Caldeira, A., and Coelho, R. (2023). EEMD-IF based method for underwater noisy acoustic signals enhancement in time-domain. *IEEE Signal Process. Lett.* 30, 294–298. doi:10.1109/LSP.2023.3258219

- Cole, P. T. (1997). "Location of partial discharges and diagnostics of power transformers using acoustic methods," in IEE Colloquium on Condition Monitoring of Large Machines and Power Transformers (Digest No: 1997/086), London, UK, 19–19 June 1997, 8/1–8/3. doi:10.1049/ic:19970501
- Geng, Q. S., Wang, F. H., and Zhou, D. X. (2019). "Mechanical Fault diagnosis of power transformer by GFCC time-frequency map of acoustic signal and convolutional neural network," in 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 21–23 Nov. 2019, 2106–2110. doi:10.1109/iSPEC48194.2019.8975318
- Hsu, C. H., Lee, S. L., Lin, C. C., Liu, C. S., Chang, S. Y., Hsieh, M. F., et al. (2015). Reduction of vibration and sound-level for a single-phase power transformer with large capacity," in *IEEE Trans. Magnetics* 51, 8403204. doi:10.1109/TMAG.2015.2453635
- Huang, H., Ding, S., Zhao, L., Chen, L., and Gao, H. (2020). Real-time fault detection for IIoT facilities using GBRBM-based DNN. *IEEE Internet Things J.* 7 (7), 5713–5722. doi:10.1109/JIOT.2019.2948396
- Huang, H., Xue, S., Zhao, L., Wang, W., and Wu, H. (2023). Privacy-preserving Smart energy management by consumer-electronic chips and federated learning. *IEEE Trans. Consumer Electron.*, 1. doi:10.1109/TCE.2023.3343821
- Kollias, D., and Zafeiriou, S. (2021). Exploiting multi-CNN features in CNN-rnn based dimensional emotion recognition on the OMG in-the-Wild dataset. *IEEE Trans. Affect. Comput.* 12 (3), 595–606. doi:10.1109/TAFFC.2020.3014171
- Kucera, M., Gutten, M., Karman, M., and Korenciak, D. (2022). *Analysis of the condition of dry transformers by acoustic emission measurement*. Krakow, Poland: 2022 ELEKTRO, 1–4. doi:10.1109/ELEKTRO53996.2022.9803687
- Lee, T., Shi, L., Nielsen, J. K., and Christensen, M. G. (2021). Fast generation of sound zones using variable span trade-off filters in the DFT-domain. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 29, 363–378. doi:10.1109/TASLP.2020.3042701
- Liang, G., Li, S., Liu, R., Cao, J., Hao, Y., and Chen, W. (2017). "A probabilistic maintenance scheme evaluation method for transformer based on failure rate," in 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, July 21 2017 to July 23 2017, 90–93. doi:10.1109/ICISCE.2017.29
- Lu, L., Yi, Y., Huang, F., Wang, K., and Wang, Q. (2019). Integrating local CNN and global CNN for script identification in natural scene images. *IEEE Access* 7, 52669–52679. doi:10.1109/ACCESS.2019.2911964
- Okabe, G. U., and Wada, H. (2011). Partial discharge signal propagation characteristics inside the winding of oil-immersed power transformer - using the three-winding transformer model in air. *IEEE Trans. Dielectr. Electr. Insulation* 18 (6), 2024–2031. doi:10.1109/TDEI.2011.6118640
- Sithole, E., Akumu, A. O., and Irungu, G. K. (2019). "Effects of insulation barrier board on power transformer PD detection using acoustics," in 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), Bloemfontein, South Africa, 28–30 January 2019, 431–436. doi:10.1109/RoboMech.2019.8704792
- Wang, J., Ji, F., and Chen, F. (2017). "Matrix pencil based toeplitz covariance matrix reconstruction approach for correlated weak source detection," in 2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xiamen, China, 22–25 October 2017, 1–5. doi:10.1109/ICSPCC.2017.8242568
- Wang, X., Wang, X., and Liu, Y. (2011). "Application of multiple models BP NN weighting optimal controller in medium plate cooling process," in Proceedings of the 30th Chinese Control Conference, Yantai, China, July 22–24, 2011, 1944–1949.
- Wang, Y., Jin, H., Cai, X., Gong, P., and Jiang, X. (2021). Study on the sound radiation efficiency of a typical distribution transformer. *IEEE Access* 9, 125151–125157. doi:10.1109/access.2021.3107455
- Wardani, N. U. A., Purnomoadi, A. P., Septiani, H. I., Arifianto, I., and Cahyono, B. (2011). "Condition assessment of 500/150 kV power transformer based on condition based maintenance," in Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, Bandung, Indonesia, 17–19 July 2011, 1–4. doi:10.1109/ICEEI.2011.6021600
- Wu, X., Zuo, W., Lin, L., Jia, W., and Zhang, D. (2018). F-SVM: combination of feature transformation and SVM learning via convex relaxation. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (11), 5185–5199. doi:10.1109/TNNLS.2018.2791507
- Zhang, X., Durand, L., Senhadji, L., Lee, H. C., and Coatrieux, J. L. (1998). Time-frequency scaling transformation of the phonocardiogram based of the matching pursuit method. *IEEE Trans. Biomed. Eng.* 45 (8), 972–979. doi:10.1109/10.704866
- Zhao, L., Yang, Q., Huang, H., Guo, L., and Jiang, S. (2024). Intelligent wireless sensing driven metaverse: a survey. *Comput. Commun.* 214, 46–56. doi:10.1016/j.comcom.2023.11.024



OPEN ACCESS

EDITED BY

Rong Jia,
Xi'an University of Technology, China

REVIEWED BY

Yushuai Li,
University of Oslo, Norway
Nan Li,
Shenzhen Institute of Artificial Intelligence and
Robotics for Society, China

*CORRESPONDENCE

Jialun Sun,
✉ 202108580021223@ctgu.edu.cn

RECEIVED 30 July 2023

ACCEPTED 26 February 2024

PUBLISHED 08 March 2024

CITATION

Zou H, Sun J, Ye Z, Yang J, Yang C, Li F and
Xiong L (2024), A bolt defect detection method
for transmission lines based on
improved YOLOv5.
Front. Energy Res. 12:1269528.
doi: 10.3389/fenrg.2024.1269528

COPYRIGHT

© 2024 Zou, Sun, Ye, Yang, Yang, Li and Xiong.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

A bolt defect detection method for transmission lines based on improved YOLOv5

Hongbo Zou^{1,2}, Jialun Sun^{1,2*}, Ziyong Ye^{1,2}, Jinlong Yang^{1,2},
Changhua Yang^{1,2}, Fengyang Li^{1,2} and Li Xiong³

¹College of Electrical and New Energy Engineering, China Three Gorges University, Yichang, China,
²Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China
Three Gorges University, Yichang, China, ³Power Dispatching and Control Center of Guangxi Power Grid
Company, Nanning, China

To solve the problem of bolt defects in unmanned aerial vehicle inspection that are difficult to identify quickly and accurately, this paper proposes a defect detection method based on the improved YOLOv5 anchor mechanism. Firstly, the Normalized Wasserstein distance (NWD) evaluation metric and the Intersection over Union evaluation metric are combined, and the experiment determines the appropriate weight for this combination. This way, the sensitivity of using IoU alone to small object detection anchor box threshold changes was reduced. Furthermore, Convolutional Block Attention Module is included into the head network architecture of yolov5 in order to prioritize significant information and suppress irrelevant features. Omni-dimensional Dynamic Convolution (ODConv) is used to replace convolution in MobileNetv2. The combination module is used as the new backbone of the YOLOv5 model. It simultaneously enhances the model's capability to extract bolt defect object information, minimizes calculation requirements, and achieves lightweight detection across the entire model. Compared with the original algorithm, the model detection Accuracy Precision (AP) is increased by 30.1%, the mean Accuracy Precision is increased by 30.4%. Other evaluation metrics of the model, such as GFLOPs and Parameters, all decreased slightly. The above results show that the improved algorithm proposed in this paper greatly improves the detection accuracy of the model on the premise of ensuring that the model is as small as possible.

KEYWORDS

tiny object detection, transmission line bolt defects, evaluation metrics fusion, omnidimensional dynamic convolution, anchor-based model

1 Introduction

With the rapid development of China, electricity demand is also increasing. In the face of massive demand for power supply, the coverage of transmission lines must also be expanded (Qi, 2019; SHANG et al., 2019; Cao et al., 2021). As the main support of transmission lines, overhead transmission lines are the physical pillars of the power system (Liu Jun et al., 2020; Jain et al., 2020; LUO et al., 2021). Therefore, regular inspection of transmission line defects is an important task to ensure the stable operation of the power system. Bolts are a critical component in transmission lines as they facilitate the connection and securement of individual power components. However, due to the perennial exposure of transmission lines to the external environment, wind traction, severe weather, and other factors may cause bolt defects. Common defects are loose bolts and missing bolts. In

extreme circumstances, even a minute flaw can result in the failure of transmission lines and, in more severe cases, widespread power disruptions, posing a grave threat to the stability and security of the power grid. It is evident that investigating a object detection algorithm that accurately and rapidly identifies fastener defects is of critical practical importance (MA et al., 2019; Wang B. et al., 2021; Li et al., 2021; Tian et al., 2021; Li Junlong et al., 2022; Zhang Weixin et al., 2022).

The traditional inspection method mainly relies on manual inspection. However, manual inspections have apparent drawbacks. The manual inspection operation is intensive, lengthy, and costly. In particular, some transmission lines are located in harsh environments such as mountains, snow, and rivers. In addition, the bolt has a small object and a large number, which is not convenient for manual inspection. With the development of science and technology, UAV inspection has gradually become the primary method for detecting defects in power transmission lines (Ya-wei et al., 2006; JI, 2010; LIU et al., 2019; WANG et al., 2019). At present, most of the research on overhead transmission line inspection focuses on large and medium-sized defects such as insulators, anti-bird thorns, anti-shock hammers, and bird's nests. In reference (Wang et al., 2023), an innovation was made on the loss function, proposing a new SymLoss. By using SymLoss to optimize OPENet, a dual-head network can be used to predict anti-bird thorns and position boxes simultaneously. In reference (Zhai et al., 2022), they used 3D modeling to generate artificial samples, which supplemented the problem of insufficient training caused by a small number of data samples due to the lack of actual samples. And proposed the introduction of the GCL module into Faster R-CNN to extract geometric features of shockproof hammer defects, improving model accuracy. In reference (Zhang and He, 2022), they improved bird's nest detection by adding a Swin transformer module to the YOLOv4 backbone network and using the Caitong SimOTA sample allocation strategy. The improvements results were significant. In contrast, defect detection of bolts is more difficult. The size of the bolt defect itself is particularly small, and even a single bolt accounts for no more than 1% of the inspection image. Moreover, the inspection image also has a complex background environment, which makes it easy to block the bolt, so it is easy to miss the key features of the bolt defect in the extracted features, resulting in low defect detection accuracy. A new measurement index, NWD (Wang J. et al., 2021), is proposed to address the problem of detecting small objects by reducing the sensitivity and positive and negative sample allocation errors of traditional IoU evaluation metrics. However, in the actual detection task, it is not all small objects. Hence, this paper takes this situation into account and establishes a converged measurement index, giving full play to the advantages of two evaluation metrics. In reference (Huang et al., 2022), they proposed a detection method based on USRNet and improved YOLOv5X. Firstly, USRNet is used to super-resolve reconstruction of the transmission line inspection image to reduce the interference of the complex background of the transmission line. Secondly, the multi-module improved YOLOv5X algorithm is used for detection, which realizes the accuracy of small object detection. However, it is necessary to carry out super-resolution reconstruction and then detection. Although the accuracy is



FIGURE 1
Transmission line bolt defects.

improved, there are obvious disadvantages in speed, which does not meet the purpose of real-time detection. In reference (Zhai et al., 2023) they considered the problem of insufficient samples of insulator defects in transmission lines. In reference (Luo et al., 2023), Luo Peng et al. detected defects in ultra-small bolts. Deep extraction of valuable features of ultra-small bolts occurs through the ultra-small object perception module and local bolt detection module. And through two-stage detection, end-to-end bolt defect detection has been achieved, but only a single-level object detection label needs to be provided. This detection method greatly reduces the workload of data annotation, but due to the use of a two-stage end-to-end detection approach, the detection speed cannot meet the real-time detection requirements of unmanned aerial vehicle inspections. In reference (Zhang et al., 2023), Zhang Ke et al. proposed a bolt defect detection method that cannot be visually distinguished in transmission lines, considering the different definitions of defects for bolts at different positions. Improving the basic model with an expansion encoding module, inferring bolt position knowledge through relative position encoding of transmission line images, and ultimately obtaining model output through attribute and defect classifiers can effectively determine the attributes and defect types of bolts. Based on parallel vision theory, the prior knowledge of insulators and rule standards is fused to create artificial defect samples. Although the problem of insufficient samples is solved, after the addition of excessive artificial samples, the model is easy to extract the features of artificial samples, resulting in the difficulty of extracting the features of actual samples, and the accuracy of model detection becomes lower.

Considering the above problems, this paper studies a lightweight transmission line defect detection method based on the improved YOLOv5 anchor mechanism. By fusing NWD and IoU as a new evaluation metric, it aims to solve the problem of positive and negative sample allocation. MobileNetV2 inserted into ODConv is introduced in the backbone network, which fully excavates the feature information extracted from small object defects while lightweight modeling so that real-time detection in bolt defects of transmission lines can be realized.

2 Image characteristics and detection methods of transmission line bolts

2.1 Image characteristics of transmission line inspection bolts

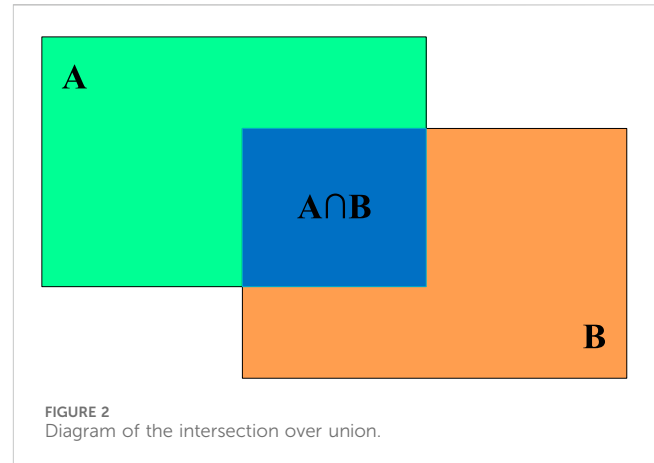
The detection object of this method is the bolt, which includes two primary defects: a missing bolt and a loose bolt. An image of the inspection bolts of a typical transmission line is shown in Figure 1. From this figure, it can be seen that in the transmission line inspection bolt image, there are the following difficulties in the inspection task:

- 1) The inspection object is tiny. When the drone is shooting, due to the distance from the object, some transmission line images have a large field of view, which is prone to a lack of appearance information such as texture, shape, and color. It often leads to failures, and small objects are difficult to detect.
- 2) The background of the detected object is complex. In the complex background of a transmission line, objects that are blurred due to image quality are difficult to detect because their contour features are not prominent and difficult to distinguish well from the environmental background. Objects also lack clearer contour features to distinguish them from occlusions, which makes object detection difficult. In view of the above characteristics of transmission line detection images, there is a need to improve the image quality of transmission line fault object detection and to enhance the ability of the object detection algorithm to detect small objects.

2.2 Transmission line inspection bolt image detection method

Aiming at the characteristics of the above-mentioned transmission line inspection bolt images, this paper proposes a new object detection idea of fusion evaluation metrics and omnidimensional dynamic convolution insertion into lightweight backbone networks. In this paper, the conventional measurement method is improved to reduce the sensitivity of small object detection tasks based on anchors and improve the ability of positive and negative sample allocation. At the same time, dynamic convolution is used to extract the fine-grained pixel characteristics of small objects and insert them into MobileNetv2 to replace convolution as the backbone network. This enhances the detection speed of the entire model.

First, improve the anchor mechanism of YOLOv5. The optimal weight of NWD and IoU fusion was determined experimentally and used as a new evaluation metric. While improving the overall accuracy of the model, it also improves the training convergence speed of the model. Second, replace the backbone part of the YOLOv5 network with mobileNetv2 plugged into ODconv as the new backbone network. While enhancing the performance of the model in identifying bolt defects in transmission lines, it also improves the detection speed of the whole model. It optimizes the performance of the entire model.



3 YOLOv5's anchor mechanism improved

The YOLOv5 model is an anchor-based model. The anchor mechanism is a technique for small objects. Because it can produce a large number of dense anchor boxes, the network can perform object classification and prediction box coordinate regression. And the dense anchor frame has a certain improvement in the recall ability of the network, which is a great advantage for small object detection. Therefore, this paper proposes a new improvement scheme for this mechanism. That is, the NWD index and the IoU index are fused as a new evaluation metric to solve the problem that positive and negative samples of object allocation in the YOLOv5 algorithm are prone to errors in small object detection to improve the performance of the model.

3.1 YOLOv5's general evaluation metric intersection over union

3.1.1 Introduction to intersection over union

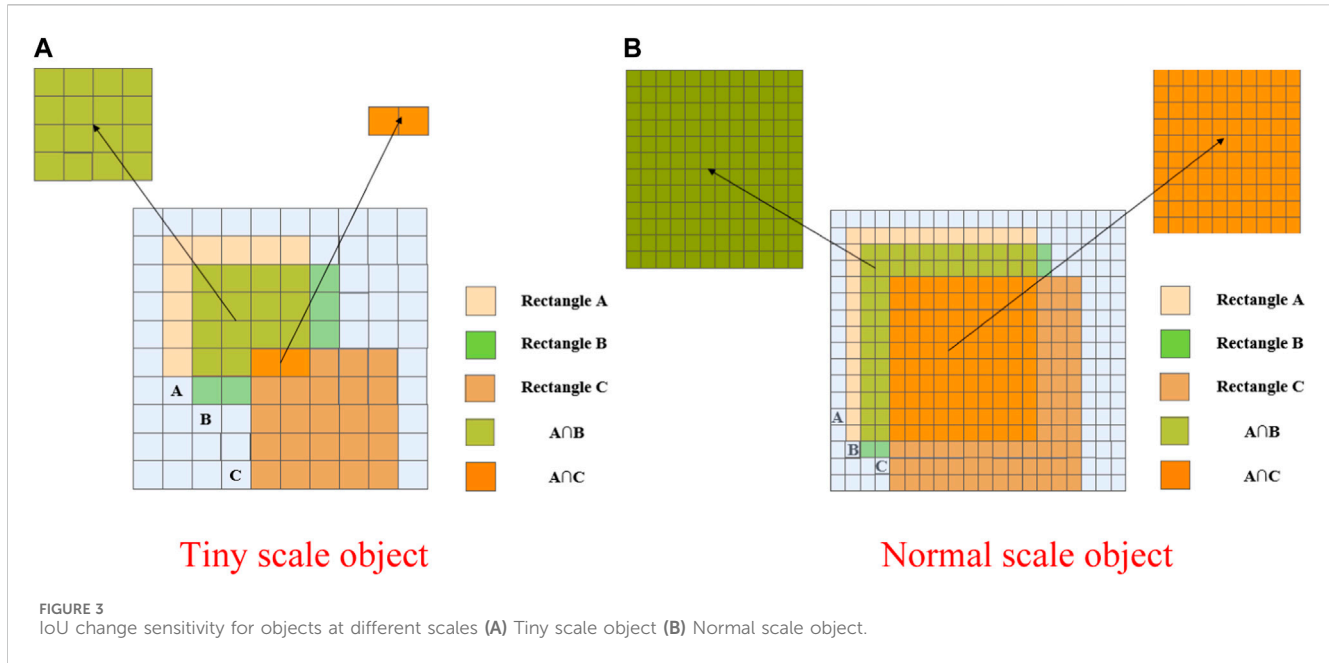
Intersection over Union is shortened to IoU. IoU is a standard that measures the accuracy of detecting objects in a specific data set. IoU is a simple measurement standard. Any task that produces bounding boxes in the output can be measured with IoU. In object recognition, a certain ratio of our commonly the prediction box to the actual box is IoU. The formula for calculating IoU is shown below:

$$IoU = \frac{A \cap B}{A \cup B} \quad (1)$$

As shown in Figure 2, IoU means the intersection of rectangle A and rectangle B divided by the union of rectangle A and rectangle B.

3.1.2 Intersection over union application scenarios

In addition to being used as an evaluation index for object detection, IoU have the following three functions under the YOLOv5 algorithm:



- 1) In the object detection of the anchor-based method, positive samples and negative samples are distinguished according to the value of the IoU.
- 2) It can be optimized directly as a loss function for bounding box regression.
- 3) Filter the prediction box in the NMS (non-maximum suppression).

3.2 Normalized Wasserstein distance

For small object objects, since most real objects cannot be standard rectangles, bounding boxes tend to have some background information. The information and background information of the object object are concentrated on the center point and boundary of the bounding box, respectively. Then, when building a two-dimensional Gaussian distribution for a bounding box, you can set the center pixel of the bounding box to the highest weight, then gradually decrease from the center point to the boundary. For a bounding box $R = (cx, cy, w, h)$, it can fit into a two-dimensional Gaussian distribution $N(\mu, \Sigma)$.

$$\text{Obey } \mu = \begin{bmatrix} cx \\ cy \end{bmatrix}, \Sigma = \begin{pmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{pmatrix} \quad (2)$$

This way, the similarity between bounding boxes is converted into a distribution distance between Gaussian distributions. where (cx, cy) is the center coordinate of the bounding box, c , w and h represents the center coordinate, width, and height respectively.

Using the Wasserstein distance to calculate the distribution distance, the second-order Wasserstein distance between different bounding boxes $\mu_1 = N(m_1, \Sigma_1)$ and $\mu_2 = N(m_2, \Sigma_2)$ is defined as:

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \|\Sigma_1^{1/2} + \Sigma_2^{1/2}\|_F^2 \quad (3)$$

Using the Gaussian distribution N_a and N_b , the formula can be simplified to:

$$W_2^2(N_a, N_b) = \left\| \left(\begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right) \right\|_2^2 \quad (4)$$

Since the result obtained above is a distance, and IoU is a ratio, the scale value is maximum 1 and minimum is 0, so this distance needs to be normalized, so the Normalized Wasserstein Distance (NWD) is finally obtained:

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right) \quad (5)$$

3.3 Construction of fusion evaluation metrics

As shown in Figure 3, little positional deviations can result in major changes to the IoU. Due to the importance of IoU in label allocation, the difference in the value size may directly lead to the allocation of negative samples that should theoretically be assigned to positive samples. When part of the object scale is too small, the degree of overlap between the anchor box and the real box will never reach the threshold. IoU itself has its flaws, which has led to the creation of many variants of evaluation metrics based on it. IoU only works when the bounding boxes overlap, so GIoU was proposed to solve this problem by adding penalties. But when the two bounding boxes contain each other, GIoU is downgraded to IoU. Subsequently, DIoU and CIoU were proposed to overcome these problems. However, GIoU, DIoU, and CIoU are all extensions of IoU, which are more commonly used in loss functions, and there is

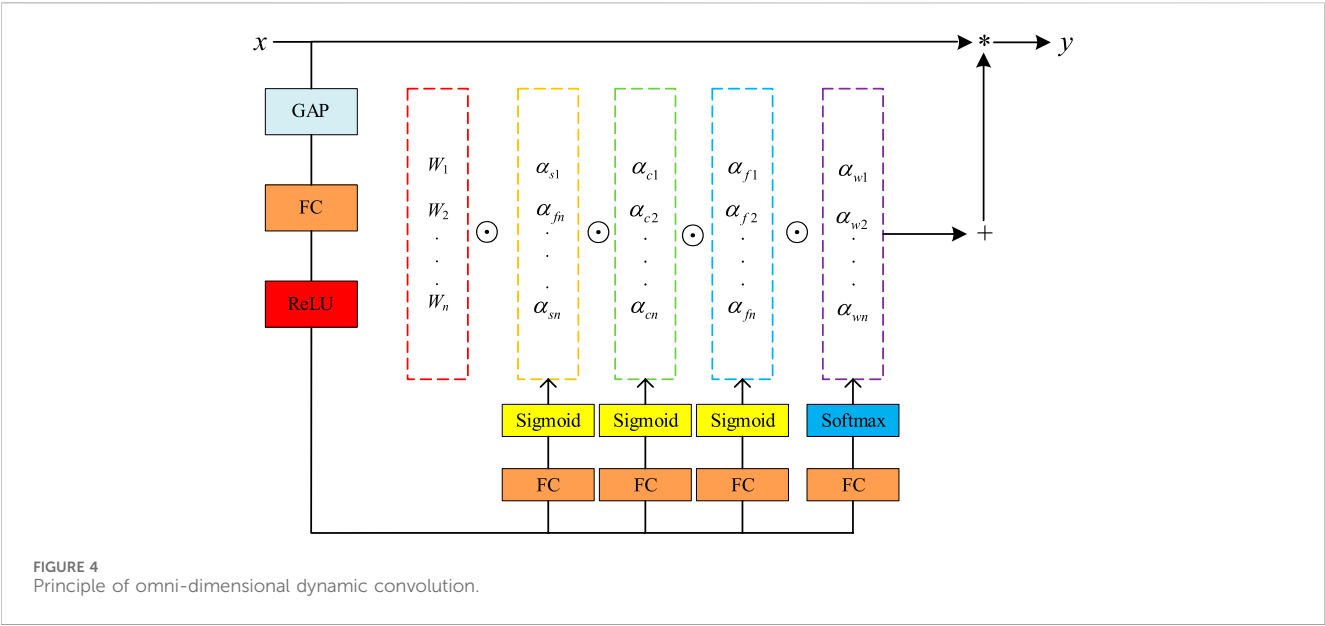
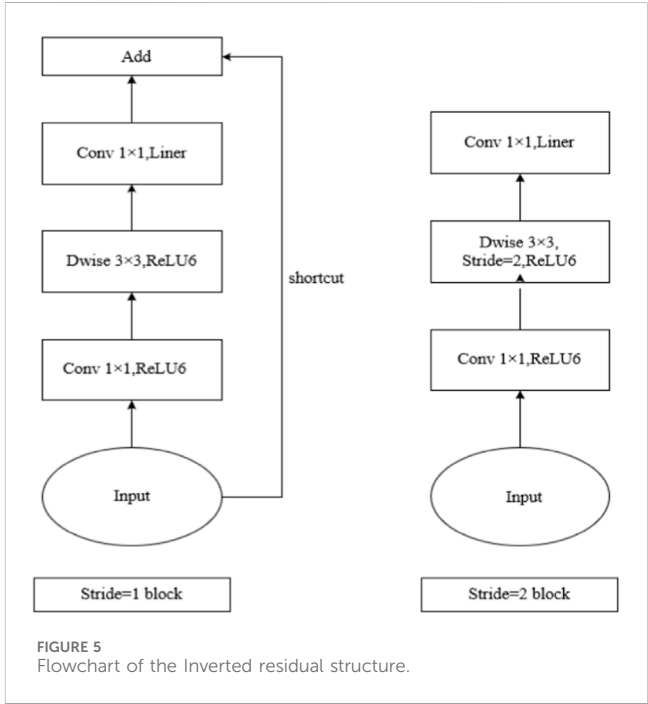


TABLE 1 MobileNetv2 network structure.

Input	Operator	t	c	n	s
$2,242 \times 3$	Conv2d		32	1	2
$1,122 \times 32$	Bottleneck	1	16	1	1
$1,122 \times 16$	Bottleneck	6	24	2	2
562×24	Bottleneck	6	32	3	2
282×32	Bottleneck	6	64	4	2
142×64	Bottleneck	6	96	3	1
142×96	Bottleneck	6	160	3	2
72×160	Bottleneck	6	320	1	1
72×320	Conv2d 1×1		1,280	1	1
$72 \times 1,280$	Avgpool 7×7			1	
$1 \times 1 \times 1,280$	Conv 1×1		k		

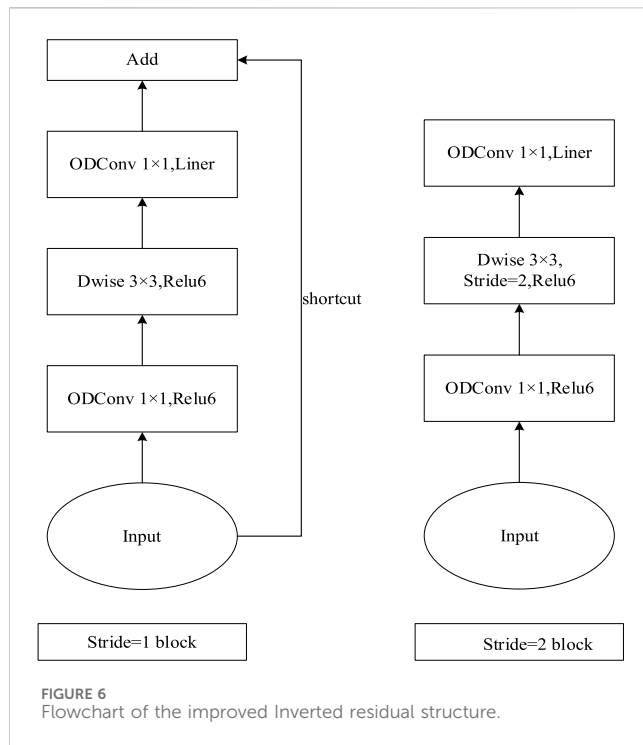
still a problem that the position offset of small objects is sensitive in label allocation. At this point, using NWD as a new evaluation metric is considered. The main step is to model the bounding box as a two-dimensional Gaussian distribution and then use NWD to measure the similarity of the derived Gaussian distribution. NWD can measure the similarity of distributions without overlap, and NWD is not sensitive to the scale of the object, making it ideal for measuring the similarity of small objects.

However, there are also problems with using NWD alone as a model evaluation metric. The use of the NWD algorithm involves the optimal transportation algorithm, which increases the running complexity of the model, makes the training speed of the model slower, and also affects the training effect. Secondly, NWD performs well in object detection tasks for very small pixel objects. The dataset in this paper has both small object defects and very small object defects, so the effect of NWD alone is not good.



To this end, this paper fuses two evaluation metrics at the same time and determines the optimal combined weights of the two through experiments. We use it as a new evaluation metric to improve model's accuracy in identifying bolt defects while also speeding up the convergence of the model.

As shown in Figure 3A shows the variation of IoU in tiny scale object; Figure 3B shows the variation of IoU in normal scale object. In Figure 3A, the value of IoU for A and B is calculated to be 0.47, and the IoU for A and C is calculated to be 0.04. In Figure 3B, the value of IoU for A and B is calculated to be 0.74, and the IoU for A and C is calculated to be 0.42. This shows that there are significant disadvantages of using IoU for tiny scale objects in object detection.



4 Improvement module introduction and improvement strategy

4.1 Introduction to omni-dimensional dynamic convolution

Omni-Dimensional Dynamic Convolution abbreviated as ODconv. Whereas conventional convolution has only one static convolution kernel and is independent of input samples, the traditional dynamic convolution kernel uses an attention mechanism to dynamically weight multiple convolution kernels and combine them. The calculation of ordinary dynamic convolution is shown in Eq. 6. ODconv focuses on all dimensions of kernel space, and its calculation is shown in Eq. 7.

$$y = (\alpha_{w1}W_1 + \dots + \alpha_{wn}W_n) * x \quad (6)$$

$$y = (\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \dots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * x \quad (7)$$

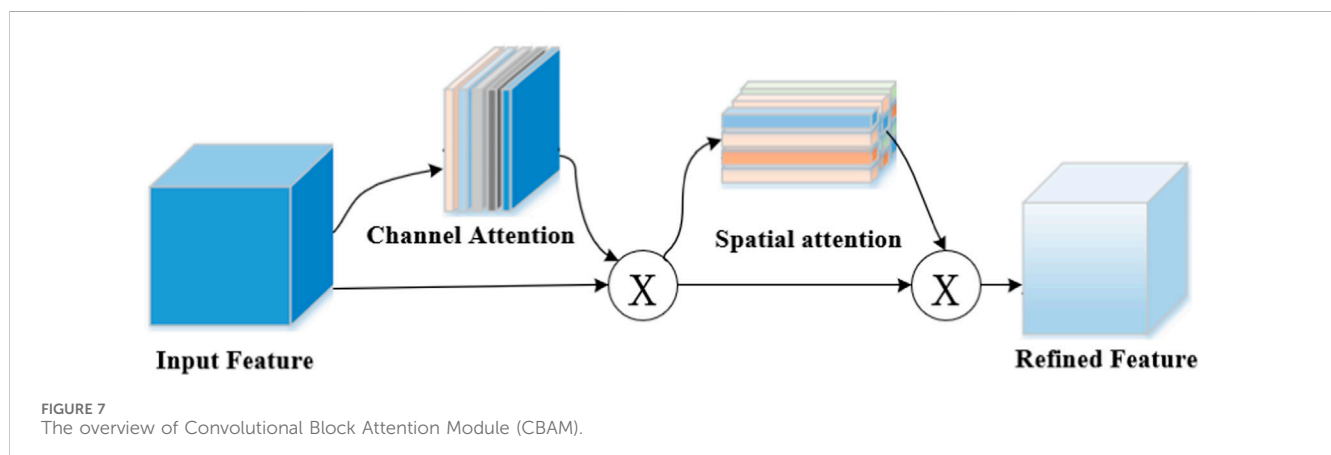
Where x represents the input feature map, y represents the output feature map, W_i represents the i th convolution kernel, α_{wi} is the attention scalar for the i th convolution kernel, and α_{si} , α_{ci} and α_{fi} represent the attention scalars along the space, input channel, and output channel, respectively. \odot represents multiplication operations in different dimensions along the kernel space.

The same calculation method as the SE attention mechanism is used. First, the input x is compressed into a feature vector of the same length as the input channel by global average pooling (GAP). Then use fully connected layers (FC) to map eigenvectors to low-dimensional space. Then go through the ReLU activation function, and generate 4 attention scalars α_{wi} , α_{ci} , α_{si} and α_{fi} of the kernel space through 4 FC layers and Sigmoid activation function or Softmax function, respectively. The process of full-dimensional dynamic convolution is shown in Figure 4.

4.2 Introduction and improvement of MobileNetv2

MobileNetv2 (Sandler et al., 2018) was proposed by the Google team in 2018. Compared with MobileNetv1, the accuracy rate is higher and the model is smaller. The structure of the MobileNetv2 network is shown in Table 1. Among them, Conv2d is a 2-dimensional convolution operation, Bottleneck is a bottleneck block composed of an Inverted residual block, Avgpool is an average pooling operation, t is the channel extension factor, c is the number of channels for the output feature matrix, and n represents the number of repeated Bottleneck and s means stride. s only represents the convolutional stride of the first Bottleneck in each convolutional stride. k represents the channel of the output vector.

The weight of many depthwise convolutions found in MobileNetv1 was found to be 0 and did not play a role. To avoid this problem, the MobileNetv2 version borrowed from the residual structure of ResNet, thereby introducing the Inverted residual. The Flowchart of the Inverted residual is shown in the figure below.



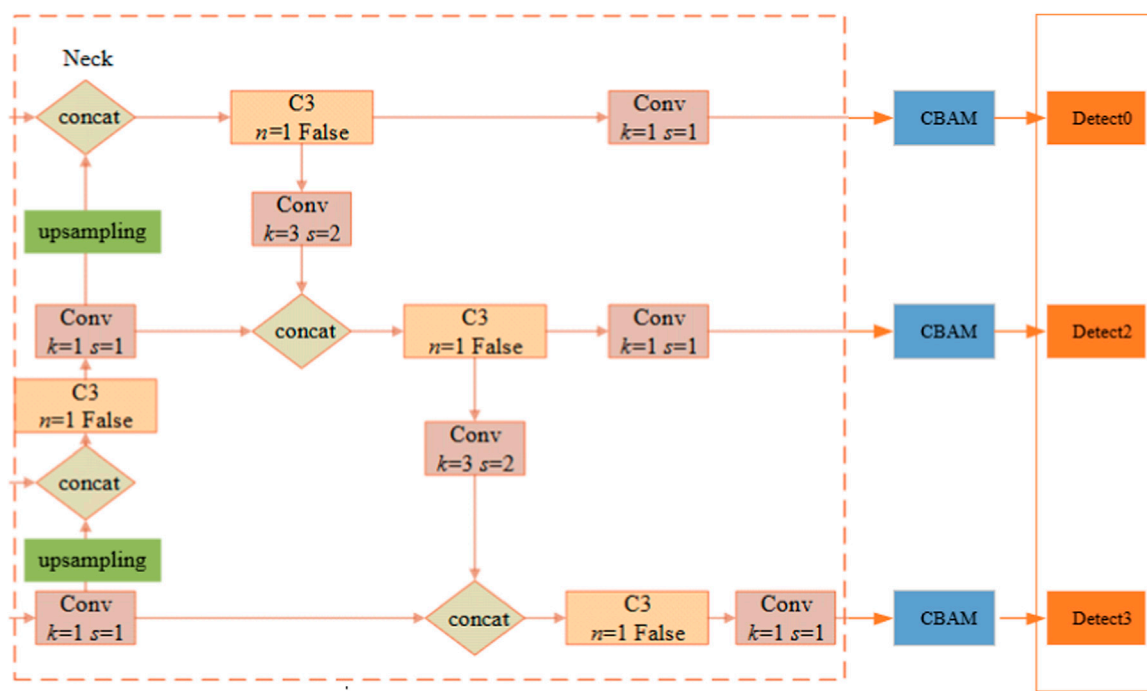


FIGURE 8
CBAM is embedded in the YOLOV5 structure diagram.

TABLE 2 Various evaluation metrics under the gradual increase in the proportion of NWD.

NWD weight	Precision	Recall	mAP
0.0	0.525	0.454	0.427
0.1	0.699	0.369	0.416
0.2	0.56	0.485	0.41
0.3	0.671	0.415	0.434
0.4	0.473	0.455	0.381
0.5	0.612	0.385	0.351
0.6	0.609	0.432	0.417
0.7	0.598	0.377	0.373
0.8	0.636	0.376	0.372
0.9	0.68	0.59	0.573
1.0	0.578	0.526	0.453

The bold values imply optimal data results.

As shown in Figure 5, Input represents the image input, Conv represents convolution operation, ReLU6 is an activation function, and its expression is $y = \text{ReLU6}(x) = \min(\max(x, 0), 6)$. Dwise represents Depthwise convolution. Liner represents a linear activation function and its expression is $y = f(x) = x$.

In the case of Stride = 1, input the image input for convolution operation, the convolution core size is 1×1 . Then use ReLU6 as the activation function for processing. Then use Dwise, and its convolution core size is 3×3 . It still used the ReLU6 function as the activation function. Immediately

then, use the convolution of the convolution core to 1×1 , and adjust the activation function to a linear activation function. When stride = 1 and the input feature matrix is the same as the output feature matrix shape, the shortcut connection operation is performed.

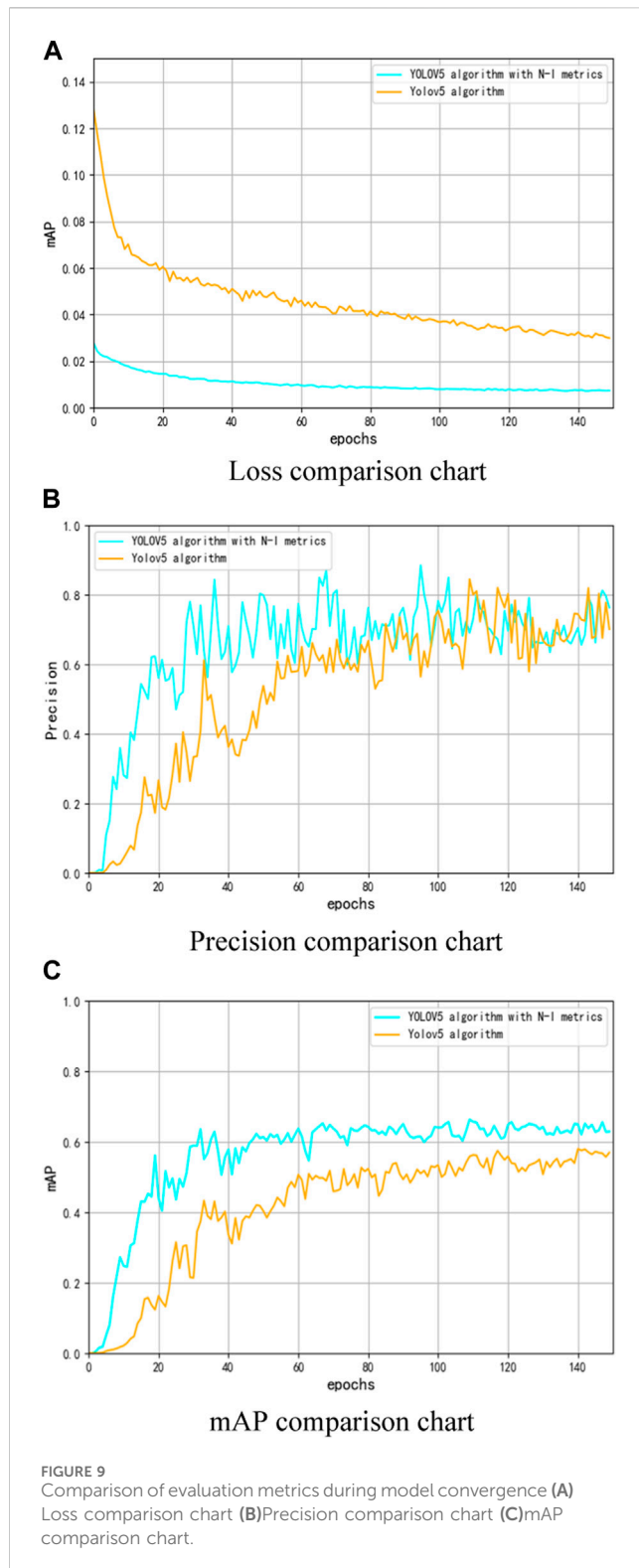
When Stride = 2, it is basically similar to Stride = 1, but there is no shortcut.

This paper uses ODConv to replace the Conv in the above structure. The improved flowchart is shown in Figure 6. The new ODConv is used and nothing else has changed.

4.3 Introduction and improvement of Convolutional Block Attention Module (CBAM)

The mechanism flow diagram of CBAM is shown in Figure 7. Firstly, perform global maximum pooling and global average pooling on the input feature map, and compress the feature map based on two dimensions to obtain two feature descriptions with different dimensions. The pooled feature maps share a multi-layer perceptron network, first reducing dimensionality through 11 convolutions and then increasing dimensionality through 11 convolutions. Overlay two feature maps with normalize the weights of each channel in the feature map through the sigmoid activation function. Multiply the normalized weights with the input feature map.

We embed the CBAM attention mechanism in the head structure of YOLOv5 for deep feature extraction, that is, focusing on important features and suppressing unnecessary features. The CBAM embedding structure is shown in Figure 8.



5 Experimental results and analysis

5.1 Evaluation metrics

The main purpose of the method proposed in this paper is to ensure the accuracy of the model and reduce its weight. That is,

improve the detection accuracy, reduce the number of model parameters and calculations, and shorten the inference time. Parameters are used to evaluate the size of the number of parameters in the model. Giga Floating Point Operations (GFLOPs) is used to indicate the computational complexity of the network. Average Precision (AP) needs to be calculated from the accuracy and recall of the model training samples. Mean Average Accuracy (mAP) is the average of the various types of object APs detected, and the detection time is the time consumed by each image detection. The formula for calculating the above evaluation metrics are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$AP = \int_0^1 P dR \quad (10)$$

$$mAP = \frac{1}{n} \sum_{i=0}^n AP_i \quad (11)$$

$$GFLOPs = HW (C_{in}K^2 + 1)C_{out} \times 10^{-9} \quad (12)$$

Where TP is the number of positive samples that have been correctly identified, FP is the number of negative samples that were incorrectly identified as positive, FN is the positive sample size of false negatives, P represents the Precision value at Recall = R, n is the total number of classes for the training sample set and i is the number of the current category. i refers to the number of channels in which the convolutional layer inputs to the tensor, C_{out} indicates the number of output channels the number of channels in which the convolutional layer outputs to the tensor, K refers to the convolution kernel size.

5.2 The optimal weight combination of Normalized Wasserstein distance and introduction to intersection over union determines the experiment

This experiment was conducted under the Python framework, with Python version 1.5.1. Using the Windows 10 operating system. Among them, CUDA version 11.4, CUDNN version 7.6, and Python version 3.8 are used in the software environment. The hardware environment is the Windows 10 operating system, and the CPU model of the testing device is 11th Gen Intel (R) Core (TM) i5-11400 @ 2.60GHz, and the GPU model is NVIDIA GeForce RTX 3060. The data in this article is actual data obtained from a transmission line in a certain province of the State Grid of China. The number of data samples is 845.

The NWD evaluation metric has a good effect on the detection of tiny objects. But in general inspection tasks, not all objects are tiny objects. Therefore, this article takes this into account, and in the model established in this article, NWD and IoU are used at the same time, and the two are fused to build a new evaluation metric.

In this paper, the NWD weight is selected to change from 0 to 1 with a step size of 0.1. The specific effects of the fusion evaluation metrics are shown in Table 2.

As can be seen from Table 2, in the overall trend, with the increase in the weight of the NWD, Precision is gradually increasing, while the change of Recall and mAP is first decreasing and then

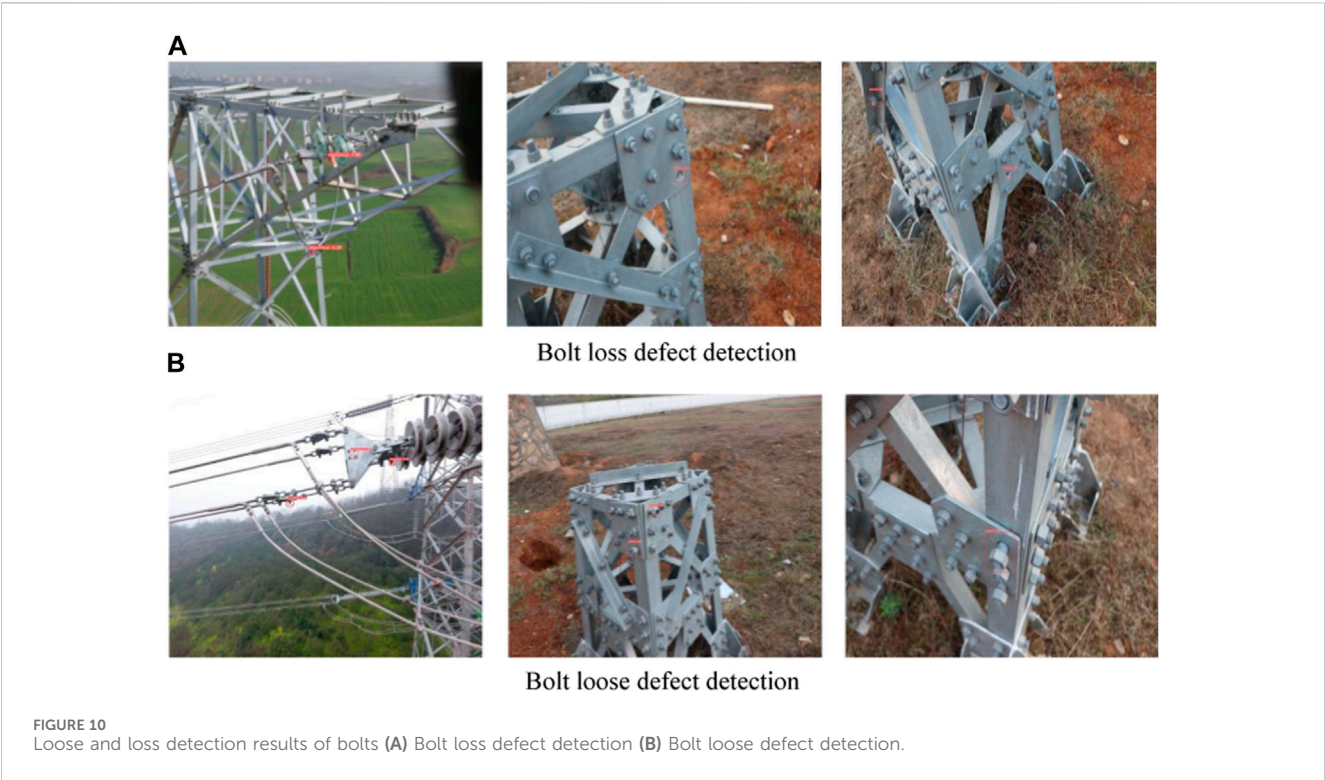


TABLE 3 Ablation experiment to improve the YOLOv5 algorithm.

	N-I metrics(0.9NWD+0.1IoU)	CBAM	MobileNetv2	ODconv	AP/ %	Map/ %	Parameter/ M	GFLOPs	Detection time/ms
1					52.25	42.7	24.3	54.3	25.8
2	√				64.0	44.6	24.3	54.3	25.9
3	√	√			64.4	46.1	28.4	62.7	29.4
4	√	√	√		65.1	47.3	25.6	59.4	27.0
5	√	√	√	√	68.0	57.3	24.9	56.9	26.2

The bold values imply optimal data results.

increasing, and finally the overall performance is better than the low-weight IoU.

The interpolation smoothing plot reflects the good characteristics of NWD in small object detection tasks as a whole. However, in Table 1, we find that the best fusion evaluation metric weight is 0.9NWD+0.1IoU. Its Precision is 0.68, Recall is 0.59, and mAP is 0.573. All evaluation metrics are optimal. 0.9NWD+0.1IoU is the optimal recombination in the detection task in this paper, and this paper calls the optimal index N-I metrics.

Therefore, this paper uses this fusion evaluation metric in the model training process. Figure 9A–C respectively reflect the changes of Loss, Precision, Recall and mAP in the training process using the N-I metrics algorithm.

As can be seen from Figure 9A, the model using the N-I metrics has basically converged at 60 epochs, and the loss value is stable at about 0.012. The model without improvement converges at 130epochs, and the loss value stabilizes at about 0.03 under

convergence. It can be seen that the use of N-I metrics can accelerate the convergence speed of the model and reduce the loss of the model.

As can be seen from Figure 9A–C, the model using the N-I evaluation metrics have good results on all evaluation metrics compared with the model algorithm without improvement.

The above experimental results show that the N-I metrics used in this paper can further mine the object semantic information and positioning information, enhance the learning ability of the network, and improve the detection performance of the network for bolt defects in transmission lines.

5.3 Ablation experiment to improve the YOLOv5 algorithm

As shown in Table 3, the ablation experimental results of the improved YOLOv5 algorithm are shown in this paper. From the

table, we can find that in the process of gradual improvement, the accuracy of the model is getting higher and higher, and the number of parameters is getting smaller and smaller. Compared with the original YOLOv5 algorithm, AP of the model is improved from 52.25% to 64% after using the N-I metrics, an increase of 22.5%. The model's mAP increased by 4.5% from 42.7% to 44.6%. It can be seen that the fusion index proposed in this paper has a large improvement in AP of bolt defect detection in transmission lines and a small improvement in mAP. It still needs improvements in other modules. After adding CBAM, AP of the improved algorithm has not changed much, and the mAP is increased by 3.4%. It is helpful to improve the performance of the entire model. After adding the CBAM module, the number of parameters increases due to the complex direction of parameter propagation. Therefore, Parameters, and GFLOPs increased slightly.

On the basis of the above, this paper uses the MobileNetv2 network as the new backbone of the YOLOv5 model, aiming to lighten the model. Neither AP nor mAP has improved much. However, compared with the YOLOv5 algorithm with the addition of CBAM, the number of Parameters of the model is reduced by 9.8%, the GFLOPs of the model are reduced by 5.2% and Detect time increased by 8.1%. After replacing ordinary convolution with ODconv, the model performance has been further improved.

In summary, the final model detection AP increases by 30.1%, mAP increases by 30.4%, the number of model Parameters, GFLOPs and Detect time have fallen in small increments compared with the original model. There are obvious improvement effects under each index, which verifies the effectiveness of the proposed model.

In summary, the algorithm proposed in this paper has good performance in various evaluation metrics. It can be used as a method to detect bolt defects in transmission lines.

6 Conclusion

Considering that the object of bolt defects of transmission lines is small, it is greatly affected by complex environmental influences, and it is difficult to detect. This paper proposes an improved YOLOv5 algorithm to solve this problem. The work done in this paper mainly has the following three aspects:

- 1) Improve the anchor mechanism of YOLOv5. YOLOv5 is the standard anchor-based model. Based on this, this article uses NWD as a new evaluation metric, but there are also problems with a single NWD evaluation metric. Therefore, this paper fuses both evaluation metrics, and determines the optimal weight of the fusion evaluation metrics through experiments. The use of N-I metrics have significantly improved both AP and mAP compared to basic algorithms. Among them, AP increased by 22.5% and mAP increased by 4.5%.
- 2) Use ODconv instead of ordinary convolution and insert it into MobileNetv2 as a new backbone. ODconv itself has a good effect in small object detection, instead of ordinary convolution, it can fully mine the semantic information of bolt defects. At the same time, in order to lightweight the model, Mobilenetv2 was used as the backbone. The organic

combination of the two also improves the performance of the model. Using MobileNetv2 without inserting ODconv, the improvement effect compared to the basic algorithm is not significant. After inserting ODconv, the model accuracy and detection speed both improved. Compared with the model using N-I evaluation metric, it improved by 4.45% on AP, 28.48% on mAP. Compared with the YOLOv5 algorithm with the addition of CBAM, the number of Parameters of the model is reduced by 12.3%, the GFLOPs of the model are reduced by 9.3% and Detect time increased by 10.9%.

- 3) The CBAM module is embedded in the head structure of yolo5. In order to focus on the important features of the bolt defect, suppress unnecessary features. In this paper, CBAM is used to focus on both channel and space. Therefore, the characteristic information of defects can be fully exploited and the model performance can be improved. Using the CBAM attention mechanism, compared to using the N-I evaluation metric, there was almost no improvement on AP, but an increase of 3.36% on mAP. After the use of CBAM, both Parameters, GFLOPs and Detect time increased slightly.

Although this article has shown significant improvement in bolt defect detection, there are still the following issues that need to be addressed:

- 1) The fusion weight of NWD and IO. In this article, the weight coefficient is used starting from 0, with a step size of 0.1, and continuing until 1. Can this problem be analyzed and solved as an optimization problem, but with optimal weights at the mathematical level.
- 2) Using MobileNetv2 on the PC end can lightweight the network. However, when chip algorithms are installed on drone onboard devices, the network actually performs poorly. How to solve the problem of ensuring that there are no network differences after transplanting PC end algorithms to drone onboard devices is also a need to consider. [(Qian et al., 2023; Matta et al., 2012; Woo et al., 2018; Li et al., 2019; Liu X. et al., 2020; Ge and Chen, 2020; Li C. et al., 2022; Zhang Y. et al., 2022; Ouyang and Yu, 2022; Huang et al., 2023)]

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JS: Conceptualization, Formal Analysis, Investigation, Software, Visualization, Writing—original draft, Writing—review and editing. HZ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Writing—original draft, Writing—review and editing. ZY: Conceptualization, Investigation, Methodology, Writing—original draft. JY: Software, Writing—review and editing. CY:

Conceptualization, Writing—original draft. FL: Writing—review and editing. LX: Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Major Science and Technology Special Project of Yunnan Provincial Science and Technology Department (202202AD080004).

Acknowledgments

The authors acknowledge the fund of Major Science and Technology Special Project of Yunnan Provincial Science and Technology Department (202202AD080004).

References

- Cao, D., Hu, W., Xu, X., Wu, Q., Huang, Q., Chen, Z., et al. (2021). Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices. *J. Mod. Power Syst. Clean Energy* 9 (5), 1101–1110. doi:10.35833/mpce.2020.000557
- Ge, Y., and Chen, S. C. (2020). Graph convolutional network for recommender systems. *Ruan Jian Xue Bao/Journal Softw.* 31 (4), 1101–1112. doi:10.13328/j.cnki.jos.005928
- Huang, Y., Liu, H., Chen, Q., Chen, Z., Zhang, J., and Yang, C. (2022). Transmission line insulator fault detection method based on USRNet and improved YOLOv5x. *Gaodianya Jishu/High Volt. Eng.* 48 (9), 3437–3446. doi:10.13336/j.1003-6520.hve.20220314
- Huang, Y., Sun, Q., Li, Y., Sun, C., and Chen, Z. (2023). Damping technique empowered robust energy flow calculation for integrated energy systems. *Appl. Energy* 343, 121168. doi:10.1016/j.apenergy.2023.121168
- Jain, P. K., Gupta, S., Bhavsar, A., Nigam, A., and Sharma, N. (2020). Localization of common carotid artery transverse section in B-mode ultrasound images using faster RCNN: a deep learning approach. *Med. Biol. Eng. Comput.* 58, 471–482. doi:10.1007/s11517-019-02099-3
- Ji, Y. (2010). The development of substation equipment inspection technology. *China Sci. Technol. Inf.* 22, 145–146.
- Li, C., Zhou, A., and Yao, A. (2022b). “Omni-dimensional dynamic convolution,” Proceedings of the ICLR 2022 - 10th international conference on learning representations. June 2022 Vienna, Austria.
- Li, J., Gu, C., Yue, X., and Li, F. (2022a). Edge-cloud computing systems for smart grid: state-of-the-art, architecture, and applications. *J. Mod. Power Syst. Clean Energy* 10 (4), 805–817. doi:10.35833/mpce.2021.000161
- Li, Y., Gao, W., Yan, W., Huang, S., Wang, R., Gevorgian, V., et al. (2021). Data-driven optimal control strategy for virtual synchronous generator via deep reinforcement learning approach. *J. Mod. Power Syst. Clean Energy* 9 (4), 919–929. doi:10.35833/mpce.2020.000267
- Li, Y., Zhang, H., Liang, X., and Huang, B. (2019). Event-triggered-based distributed cooperative energy management for multienergy systems. *IEEE Trans. Industrial Inf.* 15 (4), 2008–2022. doi:10.1109/TII.2018.2862436
- Liu, J., Jia, R., Li, W., Ma, F., Abdullah, H. M., Ma, H., et al. (2020a). High precision detection algorithm based on improved RetinaNet for defect recognition of transmission lines. *Energy Rep.* 06, 2430–2440. doi:10.1016/j.egy.2020.09.002
- Liu, X., Xu, K., Zhou, P., Zhou, D., and Zhou, Y. (2020b). Surface defect identification of aluminium strips with non-subsampled shearlet transform. *Opt. Lasers Eng.* 127, 105986. doi:10.1016/j.optlaseng.2019.105986
- Liu, J., Zhong, L., and Dong, Na (2019). Algorithm research of visual accurate alignment for substation inspection robot. *Industrial Instrum. Automation* 6, 8–13. doi:10.1109/POWERCON.2018.8602257
- Luo, P., Wang, B., Wang, H., Ma, F., Ma, H., and Wang, L. (2023). An ultrasmall bolt defect detection method for transmission line inspection. *IEEE Trans. Instrum. Meas.* 72, 1–12. doi:10.1109/TIM.2023.3241994
- Luo, P., Wang, Bo, Ma, H., Ma, F., Wang, H., and Zhu, D. (2021). Defect recognition method with low false negative rate based on combined object. *High. Volt. Eng.* 47 (02), 454–464. doi:10.13336/j.1003-6520.hve.20200701
- Ma, F., Wang, Bo, Dong, X., Wang, H., Luo, P., and Zhou, Y. (2019). Power vision edge intelligence: power depth vision acceleration technology driven by edge computing. *Power Syst. Technol.* 44 (6), 2020–2029. doi:10.1088/1742-6596/2166/1/012056
- Matta, N., Rahim-Amoud, R., Merghem-Boulahia, L., and Jrad, A. (2012). “Enhancing smart grid operation by using a WSN for substation monitoring and control,” in *Proceedings of the 2012 IFIP wireless days* Dublin, Ireland: IFIP Wireless.
- Ouyang, J., and Yu, H. (2022). Natural Language description generation method of intelligent image internet of things based on attention mechanism. *Secur. Commun. Netw.* 2022–11. doi:10.1155/2022/1200469
- Qi, R. U. (2019). Safety management of power system automation technology. *Sci. Technol. Inf.* 17 (20), 142–149. doi:10.1109/APAP47170.2019.9225130
- Qian, J., Lin, J., Bai, D., Xu, R., and Lin, H. (2023). Omni-dimensional dynamic convolution meets bottleneck transformer: a novel improved high accuracy forest fire smoke detection model. *Forests* 14 (4), 838. doi:10.3390/f14040838
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520. San Juan, PR, USA.
- Shang, C., Tian, L., Gao, H., et al. (2019). Substation inspection system. *Innovation Appl.* 32, 34–35. doi:10.2991/cas-15.2015.55
- Tian, G., Gu, Y., Shi, Di, Fu, J., Yu, Z., and Zhou, Q. (2021). Neural-network-based power system state estimation with extended observability. *J. Mod. Power Syst. Clean Energy* 9 (5), 1043–1053. doi:10.35833/mpce.2020.000362
- Wang, B., Ma, F., Ge, L., Ma, H., Wang, H., and Mohamed, M. A. (2021a). Icing-EdgeNet: a pruning lightweight edge intelligent method of discriminative driving channel for ice thickness of transmission lines. *IEEE Trans. Instrum. Meas.* 70, 1–12. doi:10.1109/tim.2020.3018831
- Wang, J., Xu, C., Yang, W., and Yu, L. (2021b). A normalized Gaussian Wasserstein distance for tiny object detection. <https://arxiv.org/abs/2110.13389>.
- Wang, Z., Liu, J., Wang, F., and Wu, J. (2023). OPENnet: object position embedding network for locating anti-bird thorn of high-speed railway. *IEICE Trans. Inf. Syst.* 52, 824–828. doi:10.1587/transinf.2022DLL0011
- Wang, Bo, Ma, F., Dong, X., Wang, P., Ma, H., and Wang, H. (2019). Electric power depth vision: basic concepts, key technologies and application scenarios. *Guangdong Electr. Power* 32 (9), 3–10.
- Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). CBAM: convolutional block attention module. <https://arxiv.org/abs/1807.06521>.

Conflict of interest

Author LX was employed by Power Dispatching and Control Center of Guangxi Power Grid Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Ya-wei, B. A. I., Zheng, Y.-feng, Guo, F., Guo, H.-de, Yang, H., and Wang, Y. (2006). Substation equipments inspection and defect management system based on centralization control pattern. *Power Syst. Technol.* 51, 186–188.
- Zhai, Y., Yang, K., Wang, Q., and Wang, Y. (2023). Disc insulator defect detection based on mixed sample transfer learning. *Zhongguo Dianji Gongcheng Xuebao/Proceedings Chin. Soc. Electr. Eng.* 43 (7), 2867–2876. doi:10.13334/j.0258-8013.pcsee.212774
- Zhai, Y., Yang, K., Zhao, Z., Wang, Q., and Bai, K. (2022). Geometric characteristic learning R-CNN for shockproof hammer defect detection. *Eng. Appl. Artif. Intell.* 116, 105429. doi:10.1016/j.engappai.2022.105429
- Zhang, K., Lou, W., Wang, J., Zhou, R., Guo, X., Xiao, Y., et al. (2023). PA-DETR: end-to-end visually indistinguishable bolt defects detection method based on transmission line knowledge reasoning. *IEEE Trans. Instrum. Meas.* 72, 1–14. doi:10.1109/TIM.2023.3282302
- Zhang, W., Hu, Bo, Xie, K., Shao, C., Niu, T., Yan, J., et al. (2022a). Short-term transmission maintenance scheduling considering network topology optimization. *J. Mod. Power Syst. Clean Energy* 10 (4), 883–893. doi:10.35833/mpce.2020.000937
- Zhang, Y., Wang, B., Ma, F., Luo, P., Zhang, J., and Li, Y. (2022b). High-precision detection method of irregular outer surface defects of power equipment based on domain adaptation network. *Gaodianya Jishu/High Volt. Eng.* 48 (11), 4516–4526. doi:10.13336/j.1003-6520.hve.20220047
- Zhang, Z., and He, G. (2022). Recognition of bird nests on power transmission lines in aerial images based on improved YOLOv4. *Front. Energy Res.* 10. doi:10.3389/fenrg.2022.870253



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Linwei Sang,
University of California, Berkeley, United States
Bi Fan,
Shenzhen University, China

*CORRESPONDENCE

Wei Li,
✉ 101012237@seu.edu.cn

RECEIVED 30 January 2024

ACCEPTED 06 March 2024

PUBLISHED 20 March 2024

CITATION

Li W, Zhao W, Li J, Li J and Zhao Y (2024),
Abnormal line loss identification and category
classification of distribution networks based on
semi-supervised learning and
hierarchical classification.
Front. Energy Res. 12:1378722.
doi: 10.3389/fenrg.2024.1378722

COPYRIGHT

© 2024 Li, Zhao, Li, Li and Zhao. This is an open-
access article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Abnormal line loss identification and category classification of distribution networks based on semi-supervised learning and hierarchical classification

Wei Li^{1,2*}, Wen Zhao¹, Junmin Li¹, Jie Li¹ and Yankai Zhao¹

¹Lvliang Power Supply Company, State Grid Shanxi Electric Power Company, Lvliang, China, ²School of Electrical Engineering, Southeast University, Nanjing, China

Line loss refers to the electrical energy that is dissipated as heat during the transmission and distribution of electricity through power lines. However, unusual causes, such as grid topology mismatch and communication failure, can cause abnormal line loss. Efficient abnormal line loss detection contributes not only to minimizing energy wastage and reducing carbon emissions but also to maintaining the stability and reliability of the entire distribution network. In actual situations, the cause of abnormal line loss is not labeled due to the expensive labor cost. This paper proposes a hierarchical abnormal line loss identification and category classification model, considering the unlabeled and unbalanced sample problem. First, an abnormal line loss identification model-based random forest is established to detect whether the line loss is abnormal. Then, an abnormal line loss category classification model is developed with semi-supervised learning for line loss abnormal category classification, considering the unlabeled samples. The real dataset in China is utilized to validate the performance of the proposed model. Its reliability implies the potential to be applied to real-world scenarios to improve the management level and safety of the power grid.

KEYWORDS

distribution network, line loss, reasoning analysis, semi-supervised learning, XGBoost, random forest

1 Introduction

The line loss rate is an essential indicator of economy and technology in the low-voltage distribution network (DN) (Sayed and Takeshita, 2011; Luo et al., 2021; Sun et al., 2022). With access to distributed generation and flexible load, DN becomes increasingly complex. Meanwhile, with the increasing electricity demand, a certain quantity of line loss in DN is generated. However, limited by the metering accuracy of data acquisition devices and the reliability of transmission systems, line loss identification in DN is usually completed by labor (Jing et al., 2019). Due to the incomplete installation of metering instruments of low-voltage substations and customers (Zhu and Lin, 2021; Raghuvamsi et al., 2022), it is challenging to analyze the causes of line loss.

With the establishment of big data centers and the development of machine learning, power supply corporations have gradually started to analyze line loss based on data-driven models to improve the economic benefits. According to the data source, the data-driven line loss analysis can be divided into user-oriented data analysis and DN substation area data

analysis. Gunturi and Sarkar (2021) proposed an electricity theft detection model based on an ensemble machine learning model. The model applied the statistical method of oversampling to solve the over-fitting problem during the training process. Based on the terminal acquisition data, the model could identify the line loss anomaly in a small-scale DN. Buzau et al. (2020) used a user-side line loss identification algorithm based on a hybrid depth neural network to detect non-technical losses. The algorithm integrated a long short-term memory network and a multi-layer sensing machine, which were used for processing the original data and integrating non-time series data. Chen J. D. et al. (2023) established an electricity theft detection model based on a one-dimensional convolutional neural network. It analyzed the non-technical line loss on the user side according to the complete terminal data. The above three methods (Buzau et al., 2020; Gunturi and Sarkar, 2021; Chen J. D. et al., 2023) show a significant role in line loss identification on the user side. However, they are sensitive to the quality of user-side power consumption data and lack of universality.

Regarding line loss identification in DN, a feeder loss estimation method based on the boost k -means model was developed (Chen J. et al., 2023). The analysis index system for line loss was established, and the multi-information index was calculated according to the time series data. The established characteristic indexes were imported into the boost k -means algorithm for clustering calculation, and the outliers were marked as line loss data. Wu et al. (2019) introduced an algorithm of non-technical line loss of DN identification with large samples. Based on the robust neural network model, the proposed method employed an automatic denoising encoder to pre-process data. The RNN model classified the operation data and identified the non-technical line loss value. Yao et al. (2019) analyzed the topology of a low-voltage DN and used the GBDT model to predict the abnormal line loss nodes in the substation area. Based on parameter clustering and deep learning algorithms, the parameter correlation and time series characteristics of a DN were fully considered by Liu et al. (2022) and Zhang et al. (2022). The multi-variate characteristic parameters were utilized to predict line loss events in a DN. When the topology of the DN is clear and the operation parameters are complete, identifying and predicting line loss based on the data-driven algorithm in the substation area can achieve remarkable results.

In actual operation conditions, it is difficult to accurately measure the operational parameters in the distribution network and the accuracy power consumption data (Lin and Abur, 2018; Jiang and Tang, 2020). Zhou et al. (2022) proposed a non-technical line loss identification model based on an AP reconstruction neural network. The model reconstructed and corrected the anomaly data by the AP neural network based on the simulation dataset, followed by a deep neural network to classify the data. Huang et al., (2023) constructed the electrical characteristic index system of theoretical line loss, and the power torque was proposed to identify line loss in the case of missing line data in a DN. However, this method is a supervised learning algorithm, which requires a certain amount of labeled data to train the model. In recent years, analyzing the causes of line loss has become a research focus. Power supply corporations have become interested in the causes of different line loss types. Liang et al. (2022) proposed a line loss interval calculation method based on power flow calculation and linear optimization, which was suitable for datasets with anomalies. This method fully considered

the power flow and dispatching information and analyzed the cause of area line loss. Some studies (Wang et al., 2019; Sun et al., 2023) mentioned data-driven algorithms for line loss cause analysis, locating anomalous nodes in the network topology and analyzing the abnormal causes according to parameter deviations.

With the increasing complexity of DNs, the accuracy of traditional line loss identification methods on the overall level of the DN is crucial to guarantee. All data-driven algorithms and statistical methods greatly rely on the data quality and data quantity, especially the labeled data. The unsupervised learning methods, such as the clustering algorithms, do not need the labeled data to detect the abnormal line loss. However, its performance is limited and cannot identify the abnormal category. When the abnormal line loss data occupied the main part of the whole data, the clustering algorithm would directly regard the abnormal data as the normal one. The supervised learning algorithms, such as the neural network and tree models, have a more stable performance than unsupervised learning algorithms. However, it needs enough data to support the model training to avoid the overfitting phenomenon. In the abnormal line loss detection of a DN, the labeled data are limited due to labor consumption and time cost. Thus, the performance of supervised learning used to detect abnormal line loss with limited labeled samples cannot be guaranteed. The semi-supervised learning (Van Engelen et al., 2022; Du et al., 2024) combines unsupervised learning with supervised learning. It can utilize a large amount of unlabeled data and fewer labeled data to improve model performance and achieve a better performance than supervised learning on limited labeled data.

Considering limited labeled and unbalanced sample distribution in an actual situation, this paper proposes an abnormal line loss identification and category classification based on semi-supervised learning and hierarchical classification. The main contributions of this paper are listed as follows: (1) a hierarchical framework of abnormal line loss identification and category classification is proposed, considering the unlabeled and unbalance sample problem. (2) An abnormal line loss identification model based random forest is established to identify whether substation line loss is abnormal. (3) An abnormal line loss category classification model is developed with semi-supervised learning for line loss abnormal causal reasoning, considering the unlabeled samples.

The structure of this paper is as follows: Section 1 provides an introduction and the relevant literature. Section 2 introduces the framework of the proposed hierarchical abnormal line loss identification and category classification model. Section 3 describes the details of data pre-processing and feature engineering. Sections 4 and 5 present the details of the abnormal line loss identification model and the abnormal line loss category classification model, respectively. Section 6 displays the detailed experiment results based on the real dataset. Finally, section 7 gives the conclusion.

2 The framework of the proposed model

This paper proposes an abnormal line loss identification and category classification model of a DN based on semi-supervised learning and hierarchical classification under unbalanced samples. The model is used to identify abnormal line loss in a DN and the corresponding abnormal reasons. In practical situations, there are enough labeled data for DN line loss abnormalities but few labeled

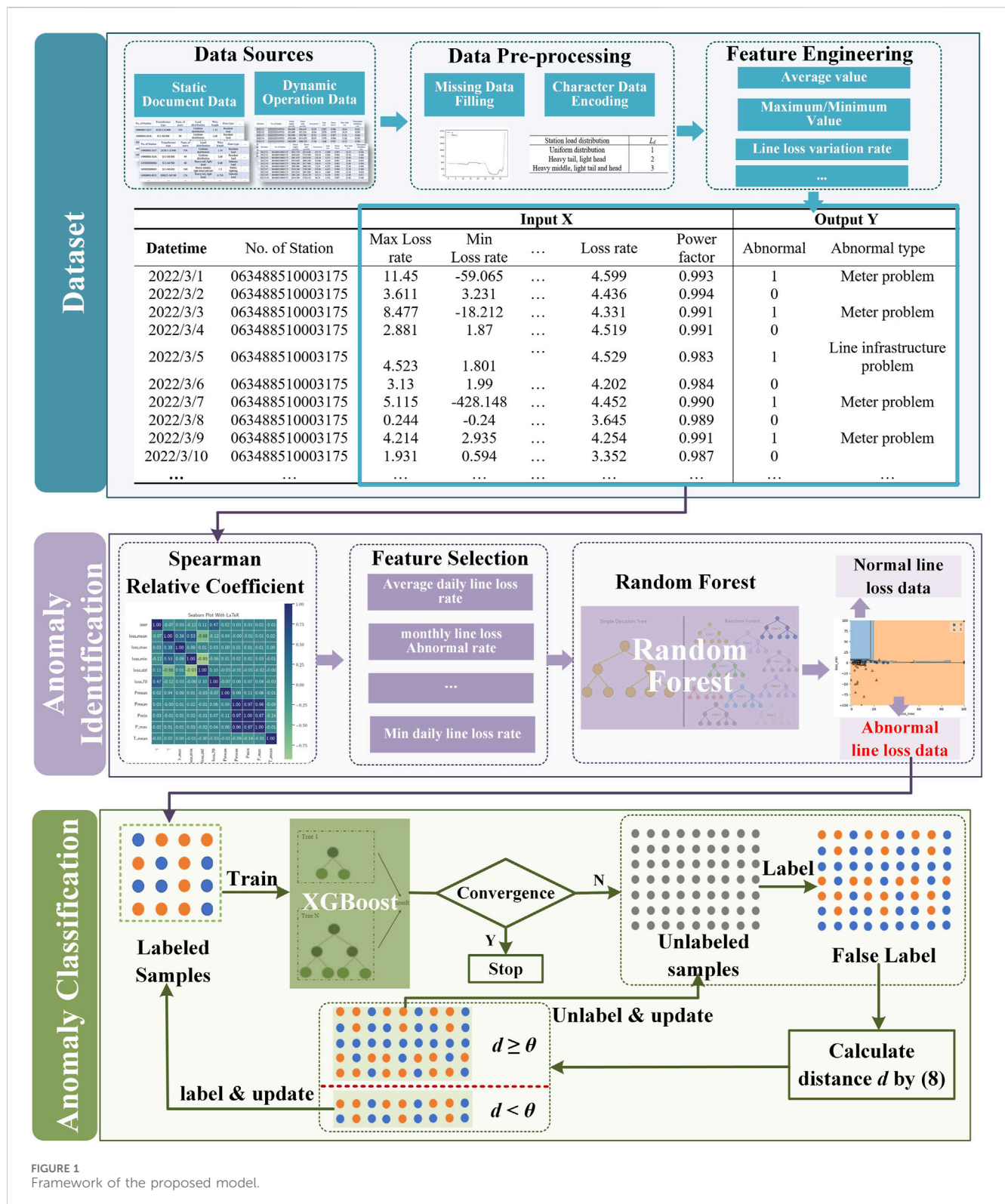


FIGURE 1
Framework of the proposed model.

data for the specific abnormal reasons. Therefore, a two-stage hierarchical classification model for identifying and reasoning abnormal line loss in a DN is proposed. In the first stage, a random forest-based abnormality identification model is established to identify whether abnormal line loss exists in the

substation. In the second stage, considering less labeled data for the specific abnormal reasons, a semi-supervised learning-based XGBoost abnormal line loss category classification model is proposed to analyze the reasons of the abnormal line loss. The overall method framework is shown in Figure 1.

TABLE 1 Examples of dynamic operation data.

Datetime	No. of substation	Power supply quantity	Power sales quantity	Line loss power	Line loss rate	Power factor	Max. Load rate	Three-phase unbalance rate
2022/3/1	063488510003175	2951.370	2815.640	135.73	4.599	0.993	24.19	0.546
2022/3/2	063488510003175	2845.620	2719.380	126.24	4.436	0.994	25.76	0.546
2022/3/3	063488510003175	2961.070	2832.830	128.24	4.331	0.991	25.09	0.504
2022/3/4	063488510003175	2935.850	2803.190	132.66	4.519	0.991	22.60	0.625
2022/3/5	063488510003175	3056.390	2917.960	138.43	4.529	0.983	25.80	0.500
2022/3/6	063488510003175	2993.900	2868.100	125.8	4.202	0.984	22.50	0.458
2022/3/7	063488510003175	2919.360	2789.380	129.98	4.452	0.990	23.50	0.540
2022/3/8	063488510003175	2913.810	2807.600	106.21	3.645	0.989	22.24	0.424
2022/3/9	063488510003175	2941.260	2816.150	125.11	4.254	0.991	22.08	0.417
2022/3/10	063488510003175	2814.540	2720.210	94.33	3.352	0.987	22.80	0.549
...

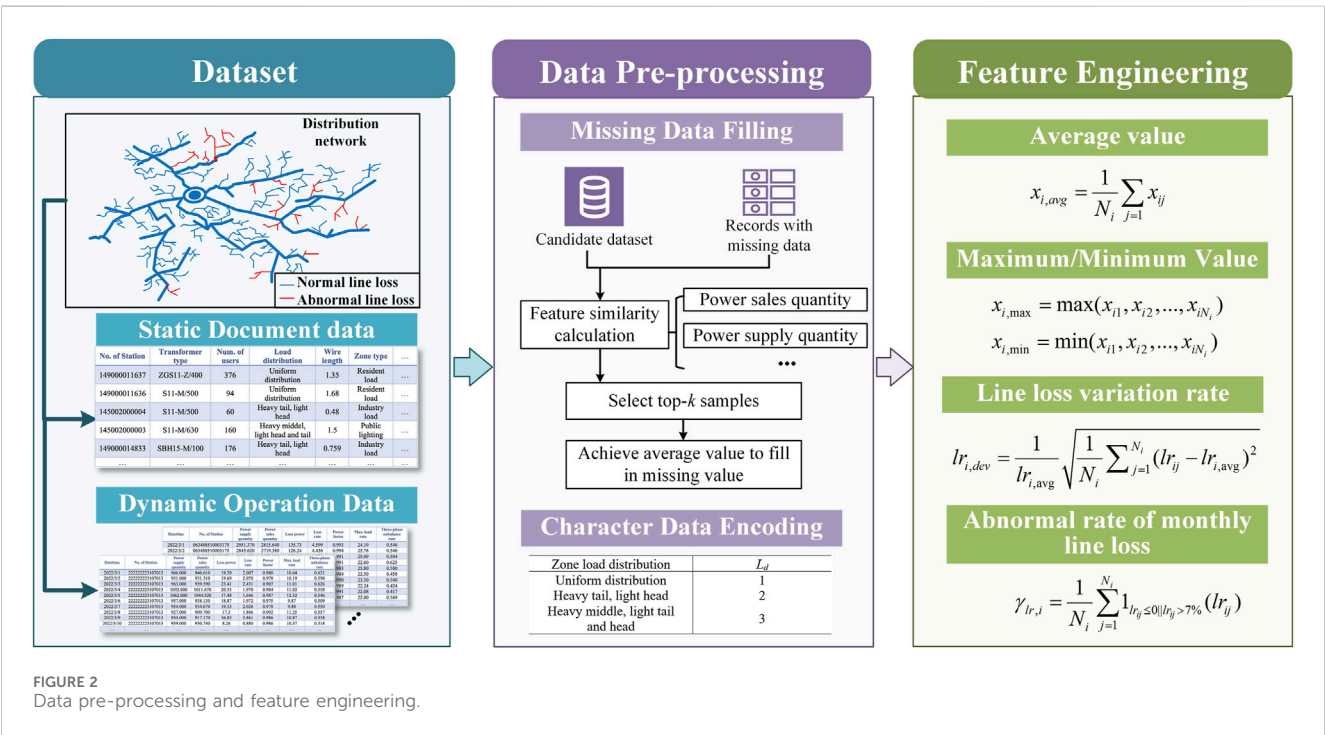


FIGURE 2 Data pre-processing and feature engineering.

- (1) Data pre-processing: the data on the distribution network substation area include static document data and dynamic operation data. In the actual data collection process, some data are missing. The *k*-nearest neighbor method is adopted to select the *k* samples that are most similar from the sample alternative set of the same substation area, and the average value of *k*-samples is taken to fill in the missing values.
- (2) Feature engineering: in the substation dynamic operation data, some features are directly related to the operation state of line loss, such as the daily line loss rate, daily maximum load rate, and daily power factor. Thus, new features are generated by the statistics of these features.
- (3) Abnormal line loss identification: the correlation analysis is carried out on all the features generated by feature engineering. The features with a high correlation coefficient are selected as the input of the abnormal line loss identification model. The dataset is divided into the training and test datasets, and the abnormal line loss identification model based on the random forest algorithm is established to identify whether the line loss is abnormal.

TABLE 2 Character data encoding of substation load distribution.

Substation load distribution	L_d
Uniform distribution	1
Heavy tail and light head	2
Heavy middle and light tail and head	3

(4) Abnormal line loss category classification: category classification is performed for the identified abnormal line losses. The common abnormal line loss causes are classified into four categories, line infrastructure problems, basic document files problem, meter problem, and theft of electricity. Considering a few data to be labeled by the abnormal category in the actual situation, a semi-supervised learning-based XGBoost abnormal line loss

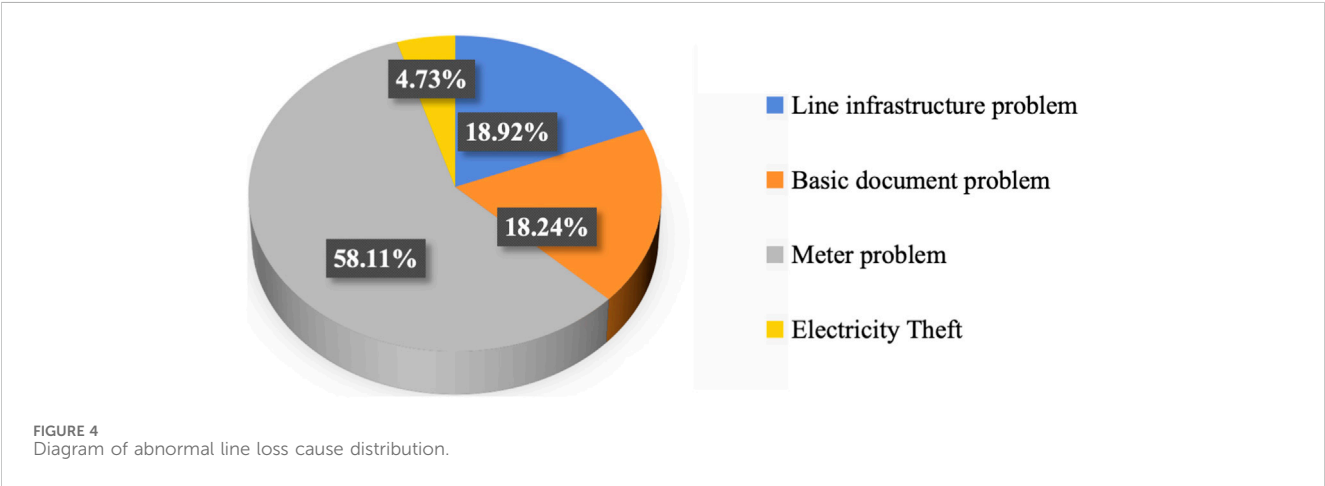
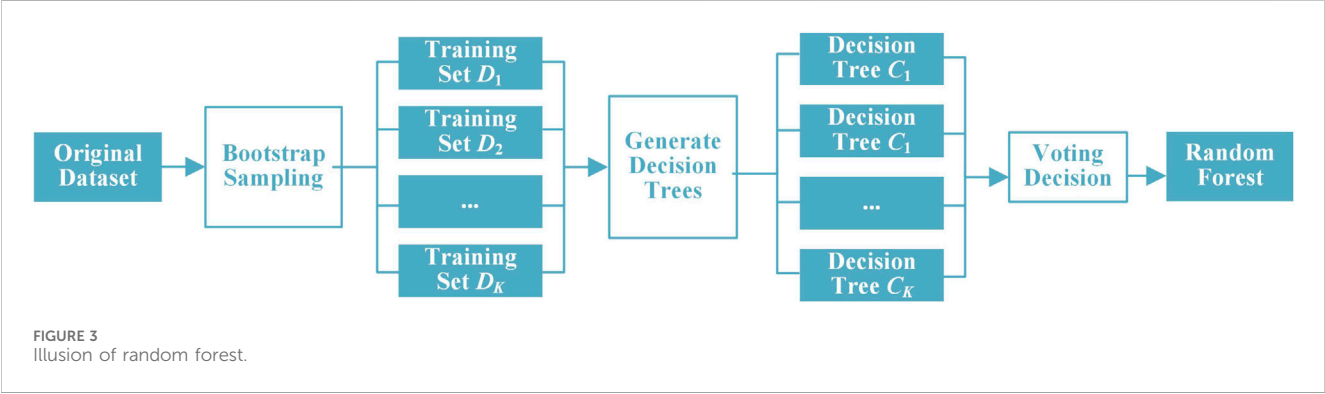


TABLE 3 Parameter settings.

Model	Parameter	Value
Random forest in abnormal line loss detection	Maximum depth of decision tree	10
	Number of decision trees	144
	Minimum number of samples in each split node	10
XGBoost in abnormal line loss category classification	Maximum depth	5
	Learning rate	0.08
	Booster	DART
	Subsample	0.75

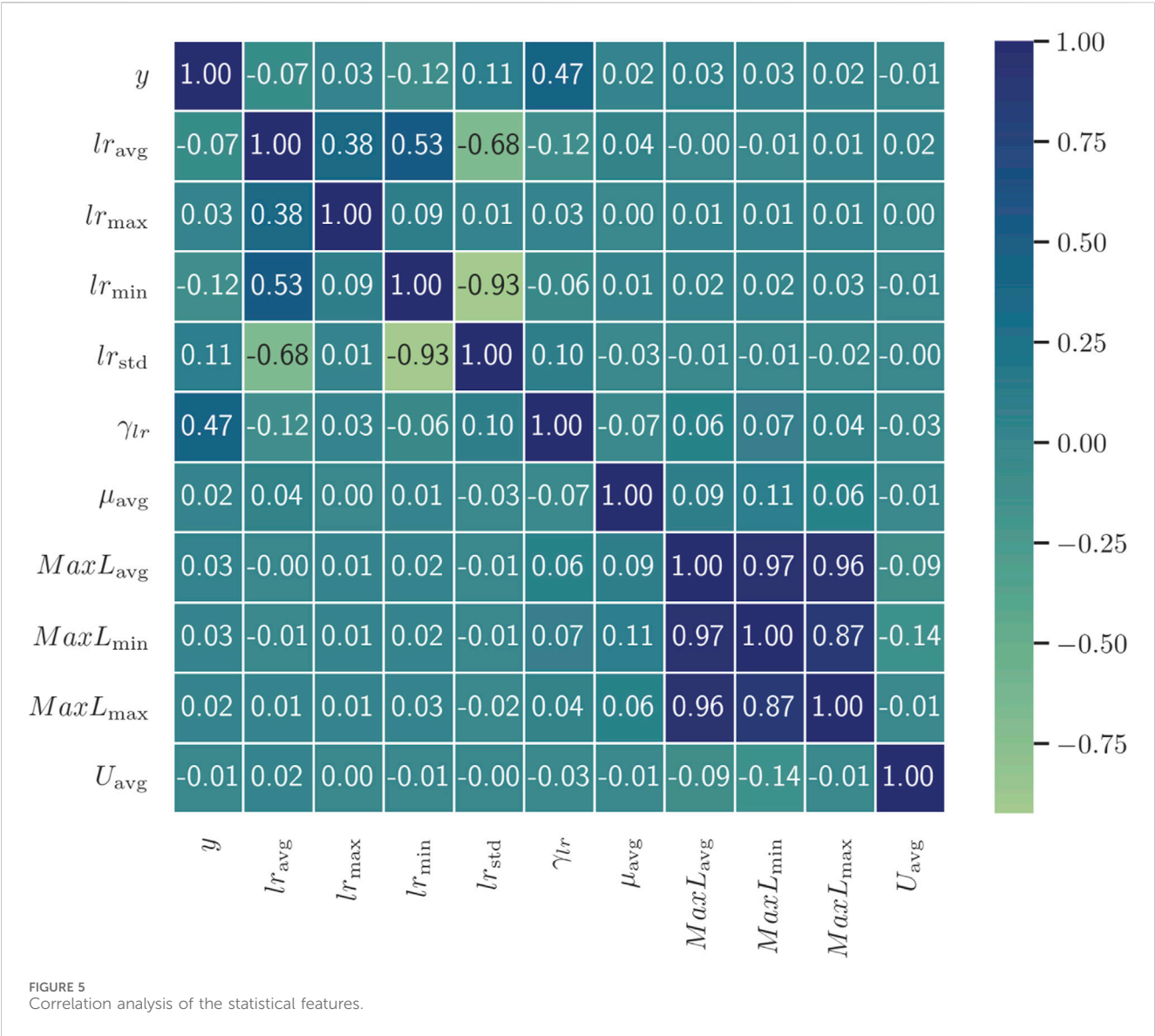


TABLE 4 Results of abnormal line loss detection and category classification.

	Abnormal line loss detection	Abnormal line loss category classification
Acc	0.9768	0.8446
P	0.9948	0.7617
R	0.9979	0.8124
F1	0.9963	0.7862

category classification model is proposed to achieve the causal reasoning analysis of the abnormal line loss.

3 Data pre-processing and feature engineering

In the process of abnormal line loss identification in a DN, the data pre-processing and feature engineering of substation operation

data are essential. By processing and extracting features from substation operation data, accurate and comprehensive features can be obtained, effectively improving the accuracy and reliability of abnormal line loss identification in a DN.

Substation data in distribution networks can be divided into two categories. One is static document data, including DN topology data, customer relationship data, the number of users, load type, transformer type, and substation load distribution. The other is dynamic operation data, including the daily input and output electricity, daily line loss rate,

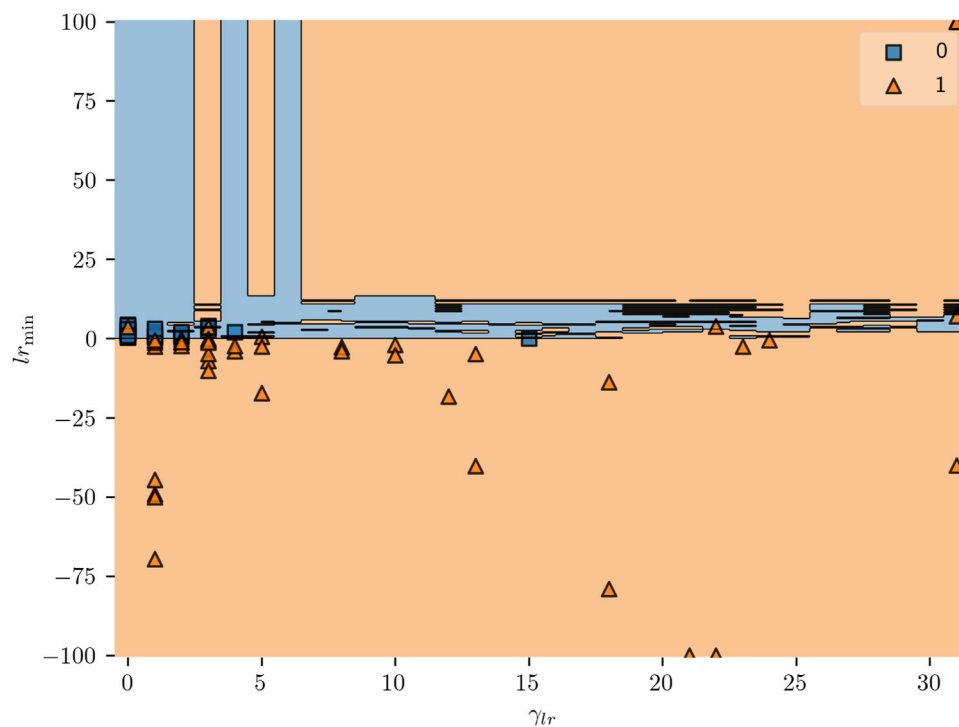


FIGURE 6
Diagram of the decision boundary of one decision tree of the random forest.

daily power factor, daily maximum load rate, daily voltage compliance rate, and daily three-phase imbalance degree, which is shown in Table 1. For dynamic document data, not only data pre-processing, such as data cleaning and completion, need to be carried out but also relevant features need to be extracted. For example, statistical measures of line loss rate, such as the average value, maximum, minimum, and variance, are significantly related to the state of line loss. The overall data processing and feature engineering processes are shown in Figure 2.

3.1 Data pre-processing

Data pre-processing mainly includes filling in missing values and encoding character data.

(1) Character data encoding

Character data encoding is carried out for the load distribution L_d and abnormal line loss categories in the substation area. Table 2 shows character data encoding for the load distribution in the substation area.

(2) Missing data filling

Upon the analysis of existing data, there were some missing data such as the daily power factor and daily three-phase unbalance in some substations. To solve this problem, the candidate set is generated by the substation. Then, the k -nearest neighbor method is adopted to select the k -samples which are the most similar from the candidate set and fill in the missing values by taking the average value of k -samples.

3.2 Feature engineering

According to the substation operation data, feature extraction is carried out on the daily power supply quantity, daily power sales quantity, daily line loss rate, daily power factor, and other data. The statistical features such as the average value, maximum value, minimum value, and variance in monthly are generated.

(1) Monthly average value

For the loss rate, l_r ; power factor, μ ; maximum load rate, $MaxL$; and the three-phase voltage unbalance rate, U , the average value is calculated with the month as the statistical length, as shown in Equation 1:

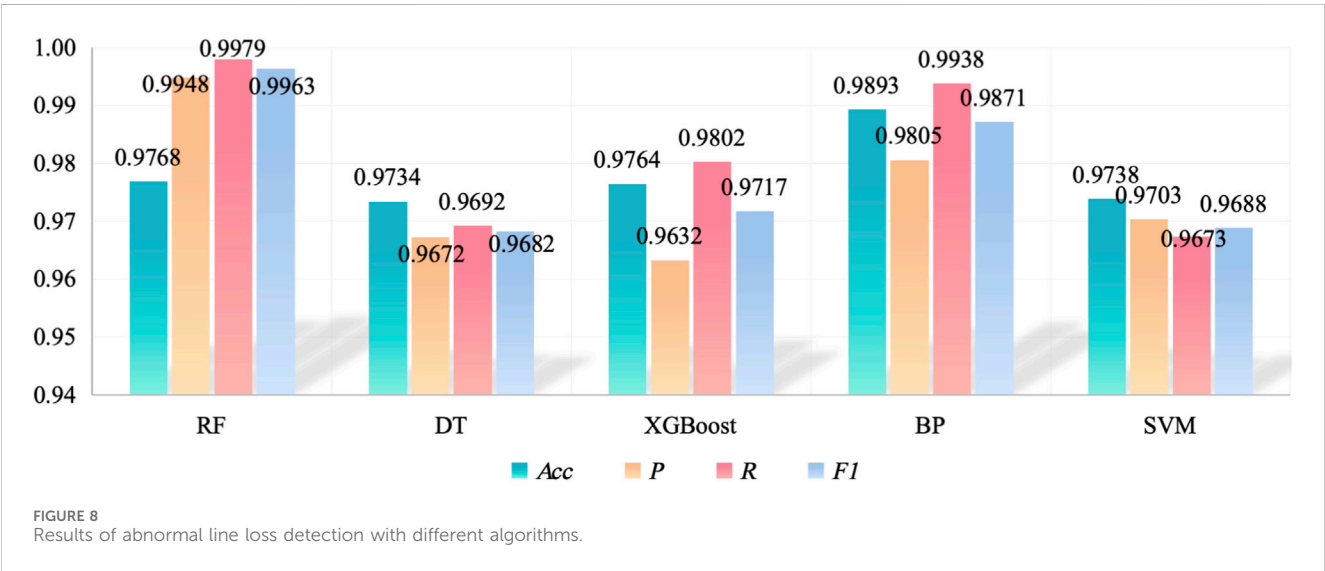
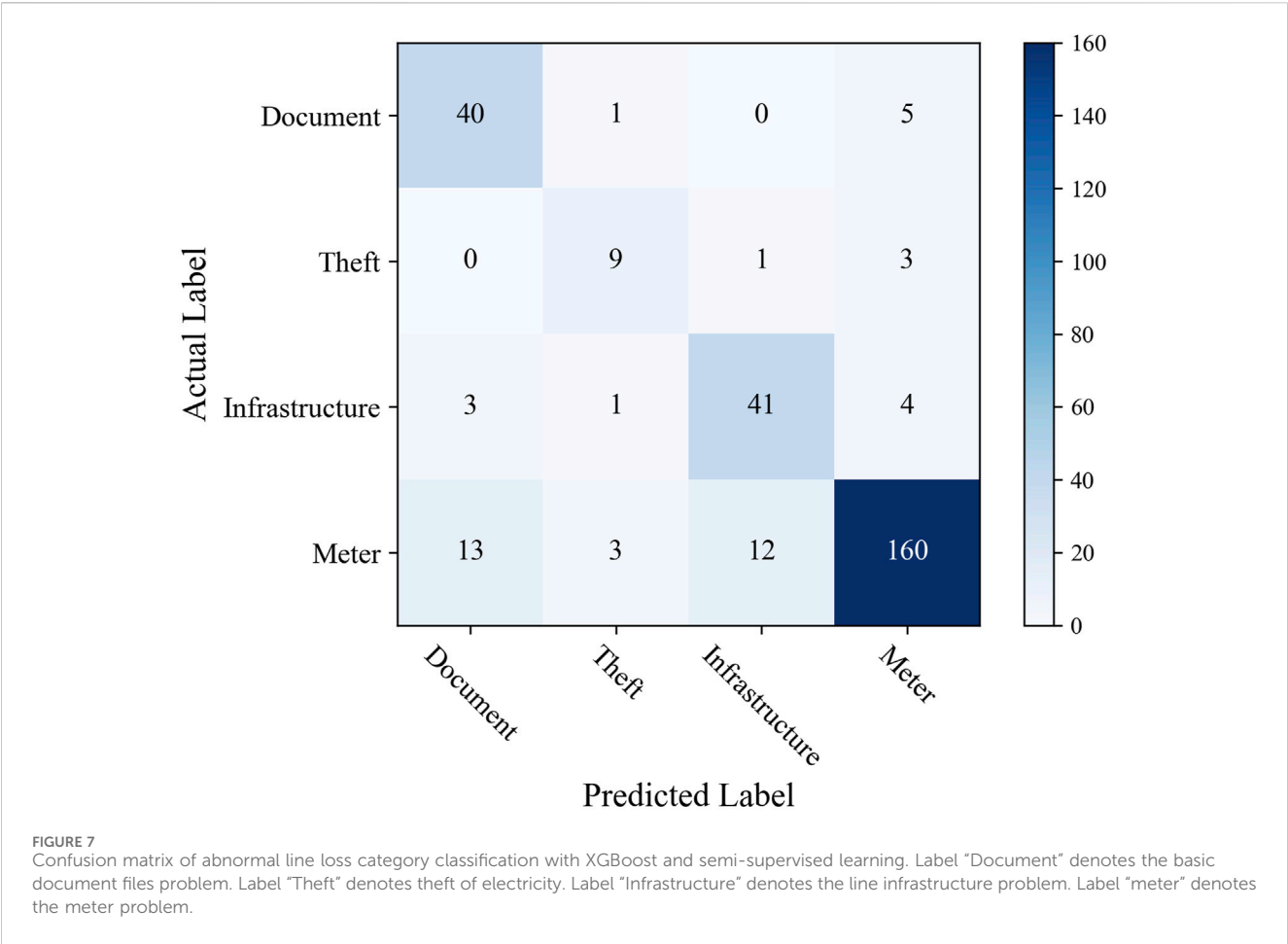
$$x_{i,avg} = \frac{1}{N_i} \sum_{j=1} x_{ij}, \quad (1)$$

where x_{ij} indicates the measured value of the j -th day of the i -th month; $x = l_r, \mu, MaxL$ or U ; $x_{i,avg}$ indicates the average value of the indicator in the i -th month; and N_i indicates the total number of days in the i -th month.

(2) Monthly maximum/minimum value

For the daily line loss rate l_r and daily maximum load rate $MaxL$, the maximum and minimum values are calculated with the monthly statistical length, which is defined by Equation 2.

$$\begin{aligned} x_{i,max} &= \max(x_{i1}, x_{i2}, \dots, x_{iN_i}) \\ x_{i,min} &= \min(x_{i1}, x_{i2}, \dots, x_{iN_i}), \end{aligned} \quad (2)$$



where x_{ij} represents the measured value of the corresponding index on the j -th day of the i -th month.

(3) Monthly fluctuation rate of daily line loss

The fluctuation of the monthly line loss rate can also reflect the abnormality of the line loss to a certain extent. Considering the difference of the average line loss rate in months, the fluctuation rate of the monthly line loss is defined as in Equation 3 in order to

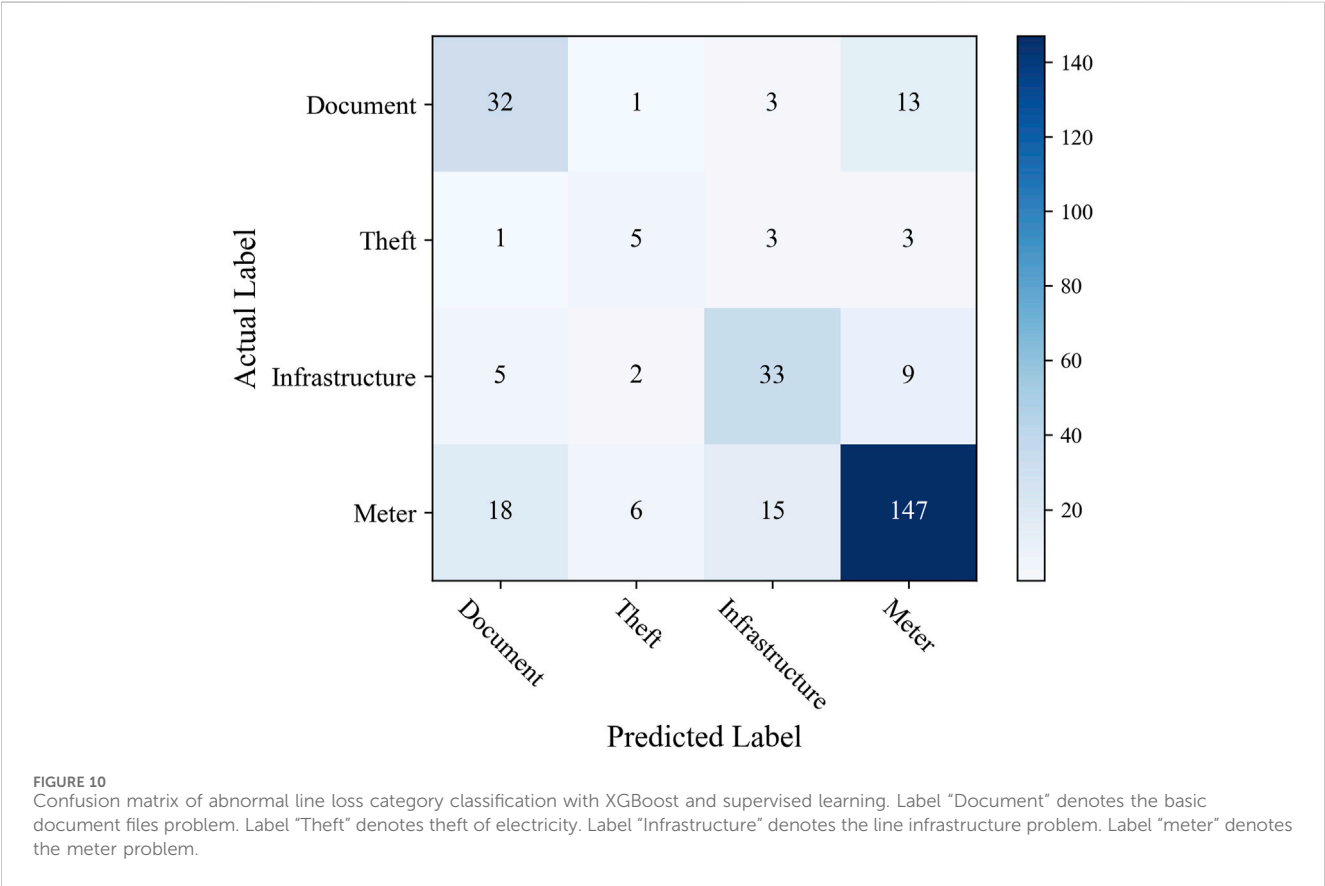
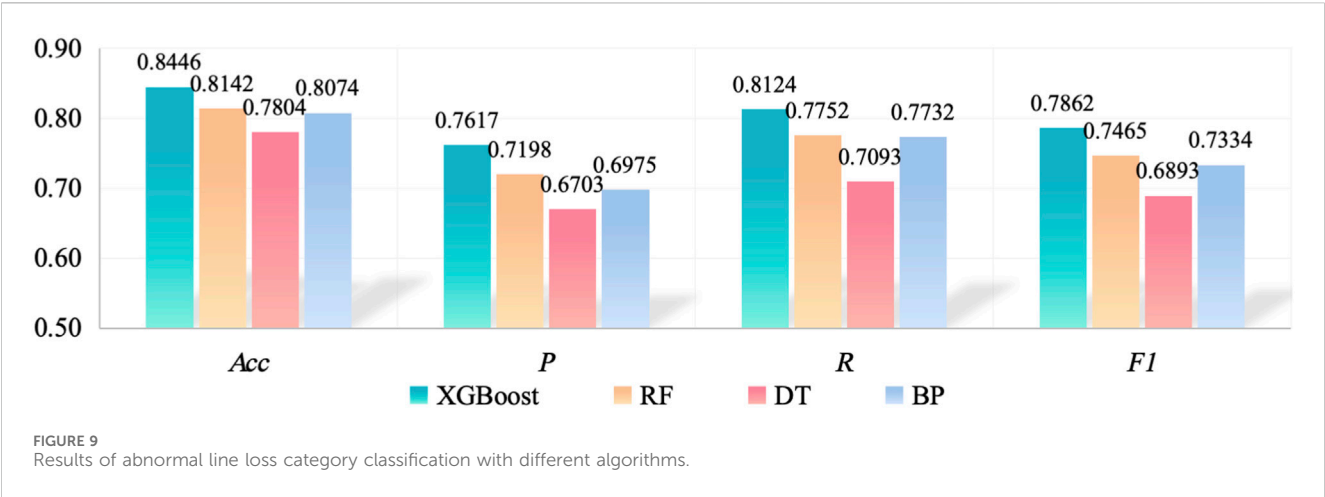


TABLE 5 Results of supervised learning and semi-supervised learning with XGBoost.

	Supervised learning	Semi-supervised learning
Acc	0.7331	0.8446
P	0.5986	0.7617
R	0.6334	0.8124
F1	0.6155	0.7862

remove the impact of the average level of the line loss rate on the statistical results.

$$lr_{i,dev} = \frac{1}{lr_{i,avg}} \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (lr_{ij} - lr_{i,avg})^2}, \quad (3)$$

where $lr_{i,avg}$ represents the average value of the line loss rate in the i -th month.

(4) Monthly abnormal rate of daily line loss

If the daily line loss rate is 0, negative, or too high, to some extent, it implies that the line loss rate may also be abnormal. Therefore, in order to reduce the influence of the accidental occurrence of the abnormal daily line loss rate, this paper defines the abnormal rate of monthly line loss, $\gamma_{lr,i}$, as shown in Equation 4. In this paper, the threshold of the excessive line loss rate is set as 7%.

$$\gamma_{lr,i} = \frac{1}{N_i} \sum_{j=1}^{N_i} 1_{lr_{ij} \leq 0 \vee lr_{ij} > 7\%}(lr_{ij}). \quad (4)$$

4 Abnormal line loss recognition based on random forest

Random forest is an inheritance algorithm based on several decision tree classifiers. The bootstrap resampling technology is used to repeatedly randomly extract parts of the samples from the original training set to form a new training set to train multiple decision trees. The final abnormal line loss identification results are obtained by combining the results of multiple independent decision trees. Compared with the single decision tree, it has higher accuracy and stability, as shown in Figure 3.

The features obtained by feature engineering are taken as the input of the random forest classifier, and the line loss abnormal is the output of the random forest classifier. Thus, the abnormal line loss identification of a DN is converted into a binary classification problem. The process is as follows:

Step 1. Dataset partitioning. The initial training set and the number of features are set. Based on the bootstrap resampling method, the samples from the original training set are repeatedly and randomly selected to form the training set D_1, \dots, D_K to build the single decision tree. The samples that have never been sampled are used to build validation datasets to estimate the performance of the model.

Step 2. Construction of a single decision tree. When constructing a single decision tree, each node is split through the principle of the minimum Gini index. When the Gini index is 0, all samples in the node belong to the same category. The Gini index is calculated as in Equation 5.

$$Gini(D) = 1 - \sum_{p=1}^P (|D_p|/|D|)^2, \quad (5)$$

where $|D|$ is the number of samples in the dataset, $|D_p|$ is the number of samples belonging to class p in the set D , and P is the number of categories.

Step 3. Decision tree integration. In K decision trees, the Boyer–Moore majority vote algorithm is used to obtain the final classification result.

In the process of training the random forest, the depth of the decision tree M_b , the number of decision trees N_b , and the minimum number of samples in each split node S_i need to be determined. This paper uses grid search and cross-validation to determine the optimal hyperparameter combination.

5 Abnormal line loss category classification based on XGBoost and semi-supervised learning

To deal with the unlabeled sample problem, semi-supervised learning is employed. An initial model is first trained with labeled data and then used to predict the unlabeled samples. The labeled samples with high confidence are added to the labeled dataset and used to retrain the model to improve the classification accuracy.

5.1 XGBoost

XGBoost adopts the idea of boosting. The basic idea is to stack the base classifiers layer by layer. Each layer gives a higher weight to the misclassified samples of the previous layer when training. The XGBoost tree is constructed by extending a node into two branches, and the layers of the nodes continue to split until the entire tree is formed. Starting from the depth of the tree equal to 0, each node traverses all the features and sorts them according to the value of the feature gain function, as shown in Equation 6. In this way, all the features are sorted according to the contribution of the features to the objective function. Then, the feature is linearly scanned to determine the best segmentation point.

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_R + G_L)^2}{H_L + H_R + \lambda} \right] - \delta, \quad (6)$$

where G_L represents the cumulative sum of the first-order partial derivation of the objective function by the samples contained in the left subtree after the current node splitting. G_R represents the cumulative sum of the first-order partial derivation of the objective function by the samples contained in the right subtree after the current node splitting. H_L represents the cumulative sum of the second-order partial derivation of the objective function by the samples contained in the left subtree after the current node splitting. H_R represents the cumulative sum of the second-order derivation of the objective function of the samples contained in the right subtree after the current node splitting. λ is the regularization parameter, and δ is the threshold to control the minimum gain of the split.

5.2 Abnormal line loss type classification based on XGBoost and semi-supervised learning

Since there are less labeled data for abnormal line loss types, most abnormal line losses only mark whether there is an anomaly but do not mark the specific reason of the anomaly. Therefore, this paper adopts the self-training semi-supervised learning method to model the abnormal line loss category classification. It trains an initial model with labeled data and then uses the model to predict the unlabeled data. The data with high confidence are added to the labeled dataset and used to retrain the model. The final model is obtained by iterating the process until the converge condition is satisfied.

According to whether the abnormal line loss type is labeled, the dataset is divided into the labeled sample dataset $\mathbf{D}_L = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2,$

$y_2), \dots, (x_n, y_n)\}$ and unlabeled sample dataset D_U . The number of sample label categories is N_c . In self-training semi-supervised learning, the pseudo-label sample selection strategy is the core part of the model performance. The purpose of the pseudo-labeled sample selection strategy is to select the samples that are more likely to be correctly labeled from the unlabeled samples and add them to the labeled samples to form a new training set so as to further improve the accuracy and generalization performance of the model. If pseudo-label samples, which are falsely labeled, are added to the training set, the performance of the model may be degraded. In this paper, pseudo-label sample selection based on the Mahalanobis distance is adopted, and the process is as follows:

In the labeled sample dataset D_L , the samples are divided according to the sample category. The sample set of class m is denoted as $D_{L,m} = \{(x_i, y_i) | y_i = m\}$, $m = 1, \dots, N_c$. The average value of its feature vector is calculated based on Equation 7.

$$\bar{x}_m = \frac{1}{|D_{L,m}|} \sum_{(x_i, y_i) \in D_{L,m}} x_i. \quad (7)$$

In the unlabeled sample dataset D_U , the corresponding pseudo-labeled sample set D_P is obtained after labeling. $y_{p,j}$ is denoted as the pseudo-label of sample x_j , $x_j \in D_U$. Suppose $y_{p,j} = m$, the Mahalanobis distance between the pseudo-label sample $(x_j, y_{p,j})$ and \bar{x}_m is calculated by Equation 8.

$$d(x_j, \bar{x}_m) = \sqrt{(x_j - \bar{x}_m) C_m^{-1} (x_j - \bar{x}_m)^T}, \quad (8)$$

where C_m is the covariance matrix of $D_{L,m}$, which is shown in Equation 9.

$$C_m = \frac{1}{|D_{L,m}| - 1} \sum_{(x_i, y_i) \in D_{L,m}} (x_i - \bar{x}_m) \left(\sum_{(x_i, y_i) \in D_{L,m}} x_i - \bar{x}_m \right). \quad (9)$$

The detailed processing of abnormal line loss category classification based on semi-supervised learning and XGBoost is shown as follows:

Step 1: the XGB model, M , is built based on the dataset $D_L = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$.

Step 2: the unlabeled sample set D_U is used as the input of model M . The corresponding pseudo-label is obtained to generate the pseudo-label sample set D_P .

Step 3: the distance $d(x_j, \bar{x}_m)$ based on Equation 8 is calculated for each pseudo-label sample $(x_j, y_{p,j}) \in D_P$.

Step 4: If $d(x_j, \bar{x}_m) < \theta$, which is the threshold, it means that the pseudo-label is acceptable. The pseudo-label sample $(x_j, y_{p,j})$ is removed from D_P and added to D_L . If $d(x_j, \bar{x}_m) \geq \theta$, it means that the pseudo-label is unreliable, and x_j is still retained in the unlabeled sample set D_U .

Step 5: D_L and D_U are updated.

Step 6: Steps 1–5 are repeated until the converge condition is satisfied. The final model is used to classify the category of abnormal line loss.

6 Experiment and results

6.1 Data source and experiment settings

In this paper, the operation data on three power supply stations in Lvliang, Shanxi Province, China, spanning half a year are used for comparison experiments. The three power stations contain 1,175 10-kV substations, which mainly include residential load, industrial load, public lighting, and commercial load. The substation operation data contain the daily active power supply, reactive power supply, line loss rate, input power, output power, power factor, maximum load rate, three-phase unbalance rate, and other data on substations spanning from May 2022 to November 2022.

In the experiment, the abnormality of the substation line loss is labeled by the month. There are a total of 7,050 samples in the dataset, including 1,503 abnormal line loss samples and 5,547 normal line loss samples. Due to the limited labor, only the abnormal causes in the part of the substation are verified, which includes 988 samples, accounting for 65.73% of the whole abnormal line loss samples. The distribution of abnormal line loss causes is shown in Figure 4. The main cause of abnormal line loss is the meter problem, including data acquisition exception and meter device fault. The electricity theft accounted for the smallest proportion. A part of the reason is that the electricity theft by users is difficult to confirm in reality due to user privacy. The detailed causes of different abnormal line loss categories are shown as follows:

- Line infrastructure problem: too long supply wire or too small wire radius and aging of the line equipment.
- Basic document files problem: distributed network topology mismatch and user-zone ownership error.
- Meter problem: data collected not at the same time, meter deviation, meter device failure, and communication failure.
- Theft of electricity: illegal use of electricity.

In the abnormal line loss recognition model, the dataset is divided as 7:3, where 70% of the data comprises the training set and 30% of the data comprises the test set. The hyperparameters of the random forest and XGBoost model used in this paper are shown in Table 3.

In this paper, the abnormal line loss detection and category classification is a two-stage classification problem. Thus, the confuse matrix is used to display the result. In stage 1, the abnormal line loss identification is a binary classification problem.

In stage 2, the abnormal line loss category classification is a multi-classification task, and the evaluation metrics include accuracy, precision, recall, and the F1-score. Considering the unbalanced sample problem, this paper utilizes the macro average value, as shown in Equations 10–13. The TP is the number of the positive samples detected as positive. The TN is the number of negative samples detected as negative. The FP is the number of negative samples detected as positive. The FN is the number of positive samples detected as negative.

$$Acc = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i}, \quad (10)$$

$$P = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{TP_i}{TP_i + FP_i}, \quad (11)$$

$$R = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{TP_i}{TP_i + FN_i}, \quad (12)$$

$$F_1 = \frac{2 \times P \times R}{P + R}. \quad (13)$$

6.2 Results of abnormal line loss identification and category classification

To further analyze the performance of feature engineering, the Spearman correlation analysis is first employed to quantify the relationship between the statistical features and abnormal line loss. The result is displayed in Figure 5. It is clear that the monthly abnormal rate of daily line loss, $\gamma_{lr,p}$ is the most important feature in abnormal line loss identification. The maximum value of line loss, the minimum value of line loss, and the average value of line loss also have a certain correlation with abnormal line loss. The three-phase unbalance rate is the least related to abnormal line loss and is not regarded as the input of the identification model.

The result of our proposed abnormal line loss identification and category classification model is shown in Table 4. It shows that a good performance is achieved in abnormal line loss identification. All the evaluation metrics obtain good results. Figure 6 displays the decision boundary of one decision tree of the random forest. It is clear that all the samples with negative line loss are recognized as abnormal. The sample with a high monthly abnormal rate and high minimal line loss rate is also identified as abnormal.

In abnormal line loss category, the classification result is not better than that of abnormal line loss identification. The small sample size and unbalanced sample distribution significantly impact the precision and recall values. The confusion matrix of the XGBoost model is presented in Figure 7. The classification result of electricity theft is the worst. The meter problem classification is the best. It is because the number of electricity theft incidents is too small and impacts the model learning. All the categories are easily misidentified as meter problems, especially electricity theft. In reality, the meter problem is the most common cause of abnormal line loss, including different abnormal line loss scenarios, such as data error, communication problem, and data collection terminal fault. Thus, other causes are easily misidentified as meter problems.

6.3 Comparison experiment

In this section, the comparison experiments are conducted from different aspects, including abnormal line loss identification with different algorithms, abnormal line loss category classification with different algorithms, and comparison of supervised learning and semi-supervised learning.

1) Comparison of abnormal line loss identification with different algorithms

In abnormal line loss identification, the decision tree (DT), XGBoost, BP, and support vector machine (SVM) are utilized as

the comparison algorithms. In DT, the max depth of tree is set as 12. In XGBoost, the learning rate is 0.1 and the number of estimators is set as 100. In BP, the number of hidden layers is set as 2, with 100 neurons in each hidden layer. The kernel function of SVM is the radial basis kernel function, and the regularization parameter is 1.

The identification results of different algorithms are shown in Figure 8. Since the abnormal line loss identification problem is a relatively simple binary classification problem, all the algorithms can achieve a good performance. From the aspect of accuracy, BP achieves the best performance. The accuracy values of RF, DT, XGBoost, and SVM are close. From the aspect of all metrics, the RF performs the best. The precision, recall, and F1-score of the RF are the highest. The precision result of BP implies that the model easily launches false alarms than RF. The performance of DT and SVM is the worst. Further analyzing the result with data, it is found that the monthly line loss with the negative daily line loss rate is easily recognized as abnormal. The abnormal monthly line loss with a small and positive line loss is the most difficult to detect compared to other abnormal line loss scenarios.

2) Comparison of abnormal line loss category classification with different algorithms

In abnormal line loss category classification, random forest, DT, and BP are used as the comparison algorithms. In random forest, the number of decision trees is set as 105, and the maximum depth of the decision tree is set as 10. In DT, the maximum depth of the tree is set as 15. In BP, the number of hidden layers is set as 3, with 85 neurons in each hidden layer. All the algorithms are conducted with the semi-supervised learning.

The result of the abnormal line loss category classification is displayed in Figure 9. It is obvious that the performance of XGBoost is the best and that of DT is the worst. Due to limited samples, the accuracy of abnormal line loss category classification is not higher than that of abnormal line loss identification. In another aspect, the input feature is generated based on monthly line loss, which cannot reflect the fluctuation of the intra-day line loss rate. In particular, the theft of electricity is closely related to the intra-day line loss rate, which cannot be well-detected.

3) Supervised learning vs. semi-supervised learning

In this section, the performance of supervised learning and semi-supervised learning is compared with XGBoost in the abnormal line loss category classification task. The supervised learning directly uses 70% of the labeled samples to train XGBoost, and the rest 30% was used for the test. The evaluation metric results are displayed in Table 5. From Table 5, it is obvious that the classification results are significantly improved by semi-supervised learning, especially recall. It is implied that the phenomenon of leaking alarm is relieved. The category of theft of electricity is the most difficult to detect. It is because of the limited electricity theft samples and because electricity theft is mostly impacted by the intra-day line loss rate. The confusion matrix of supervised learning is presented in Figure 10. Compared to Figure 7, the classification accuracy of all the categories is enhanced. For the semi-supervised learning, the unlabeled samples are used, which can help the model learn to

increase the classification accuracy. However, the current data cannot reflect the situation of intra-day line loss, and the category classification performance is limited. To further improve the abnormal line category classification, detailed line loss data are needed.

7 Conclusion

Abnormal line loss identification is crucial in distribution networks to guarantee the timely and safe power supply in grid. In actual situations, the cause of abnormal line loss is not completely labeled due to the expensive labor cost. Considering the actual limited and unbalanced samples, this paper proposed a hierarchical classification framework to identify the causal reason of the abnormal line loss. An abnormal line loss identification model-based random forest was first established to identify whether substation line loss was abnormal. Based on the results of detected abnormal line loss, an abnormal line loss category classification model was developed with semi-supervised learning and XGBoost, considering the unlabeled samples. With the help of self-training semi-supervised learning, the unlabeled samples were utilized to train the classification model to relieve the over-fitting performance. Numerous experiments were conducted on the real dataset from China. The accuracy of abnormal line loss identification was more than 97%. The accuracy of abnormal line loss category classification was around 84% under semi-supervised learning. The results highlight the good performance of the proposed hierarchical learning structure to relieve the impact of the unbalance samples, which is very helpful for future application.

In the future, more detailed abnormal line loss causes can be considered. In addition, the sampling techniques to relieve the sample unbalance can be further utilized when considering the detailed abnormal line loss causes. In summary, this research highlights the application of machine learning in abnormal line loss identification and category classification, with implications for improving the management and operation of power grids.

References

- Buzau, M. M., Tejedor-Aguilera, J., Cruz-Romero, P., and Gómez-Expósito, A. (2020). Hybrid deep neural networks for detection of non-technical losses in electricity Smart meters. *IEEE Trans. Power Syst.* 35, 1254–1263. doi:10.1109/tpwrs.2019.2943115
- Chen, J., Zeb, A., Sun, Y., and Zhang, D. (2023b). A power line loss analysis method based on boost clustering. *J. Supercomput.* 79, 3210–3226. doi:10.1007/s11227-022-04777-w
- Chen, J. D., Nanekaran, Y. A., Chen, W. R., Liu, Y. J., and Zhang, D. F. (2023a). Data-driven intelligent method for detection of electricity theft. *Int. J. Electr. Power and Energy Syst.* 148, 108948. doi:10.1016/j.ijepes.2023.108948
- Du, G., Zhang, J., Zhang, N., Wu, H., Wu, P., and Li, S. (2024). Semi-supervised imbalanced multi-label classification with label propagation. *Pattern Recognit.* 150, 110358. doi:10.1016/j.patcog.2024.110358
- Gunturi, S. K., and Sarkar, D. (2021). Ensemble machine learning models for the detection of energy theft. *Electr. Power Syst. Res.* 192, 106904. doi:10.1016/j.epsr.2020.106904
- Huang, L., Zhou, G., Zhang, J., Zeng, Y., and Li, L. (2023). Calculation method of theoretical line loss in low-voltage grids based on improved random forest algorithm. *Energies* 16, 2971. doi:10.3390/en16072971
- Jiang, W., and Tang, H. B. (2020). Distribution line parameter estimation considering dynamic operating states with a probabilistic graphical model. *Int. J. Electr. Power and Energy Syst.* 121, 106133. doi:10.1016/j.ijepes.2020.106133
- Jing, T. T., Dai, L., Xi, H. J., and Hu, Y., (2019). Method for theoretical line loss calculation of 10kV distribution district based on actual electric energy of distribution transformer secondary side. *Iop Int. Conf. Civ. Archit. Disaster Prev.* 218, 1–6. doi:10.1088/1755-1315/218/1/012152
- Liang, C., Chen, C., Wang, W., Ma, X., Li, Y., and Jiang, T. (2022). Line loss interval algorithm for distribution network with DG based on linear optimization under abnormal or missing measurement data. *Energies* 15, 4158. doi:10.3390/en15114158
- Lin, Y. Z., and Abur, A. (2018). A new framework for detection and identification of network parameter errors. *IEEE Trans. Smart Grid* 9, 1698–1706. doi:10.1109/tsg.2016.2597286
- Liu, K. Y., Jia, D. L., Kang, Z. J., and Luo, L. (2022). Anomaly detection method of distribution network line loss based on hybrid clustering and LSTM. *J. Electr. Eng. Technol.* 17, 1131–1141. doi:10.1007/s42835-021-00958-4
- Luo, F. Z., Yang, X., Yao, L. Z., Zhu, L. Z., Zhao, D. W., and Qian, M. H. (2021). “Flexible load active management method in optimization operation of distribution networks,” in Proceedings of the 2021 3rd asia energy and electrical engineering symposium (AEEES 2021), Chengdu, China 413–418. doi:10.1109/AEEES51875.2021.9403086
- Raghuvamsi, Y., Teeparthi, K., and Kosana, V. (2022). A novel deep learning architecture for distribution system topology identification with missing PMU measurements. *Results Eng.* 15, 100543. doi:10.1016/j.rineng.2022.100543

Data availability statement

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

Author contributions

WeL: writing–review and editing and writing–original draft; WeZ: writing–review and editing, writing–original draft, and data curation; JuL: writing–review and editing; JiL: writing–review and editing; YZ: writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the science and technology project of State Grid Shanxi Electric Power Company, grant number 5205J0220002. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

Authors WeL, WeZ, JuL, JiL, and YZ were employed by State Grid Shanxi Electric Power Company.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Sayed, M. A., and Takeshita, T. (2011). All nodes voltage regulation and line loss minimization in loop distribution systems using UPFC. *IEEE Trans. Power Electron.* 26, 1694–1703. doi:10.1109/tpe.2010.2090048
- Sun, B., Li, Y. F., Zeng, Y., Chen, J. H., and Shi, J. D. (2022). Optimization planning method of distributed generation based on steady-state security region of distribution network. *Energy Rep.* 8, 4209–4222. doi:10.1016/j.egy.2022.03.078
- Sun, Z., Xuan, Y., Huang, Y., Cao, Z., and Zhang, J. (2023). Traceability analysis for low-voltage distribution network abnormal line loss using a data-driven power flow model. *Front. Energy Res.* 11, 832837. doi:10.3389/fenrg.2023.1272095
- Van, E., Jesper, E., and Hoos, H. H. (2022). A survey on semi-supervised learning. *Mach. Learn.* 109 (2), 373–440. doi:10.1007/s10994-019-05855-6
- Wang, W., Bai, R., He, X., Xing, Y., Zhang, H., and Liu, J. (2019). “Development of synchronous line loss analysis and diagnosis system based on arbitrary segmentation of power grid,” in Proceedings of the 2019 IEEE 4th advanced information technology, electronic and automation control conference (IAEAC), Chengdu, China 1840–1844.
- Wu, W., Cheng, L., Zhou, Y., Xu, B., Zang, H., Xu, G., et al. (2019). Benchmarking daily line loss rates of low voltage transformer regions in power grid based on robust neural network. *Appl. Sci.* 9, 5565. doi:10.3390/app9245565
- Yao, M., Zhu, Y., Li, J., Wei, H., and He, P. (2019). Research on predicting line loss rate in low voltage distribution network based on gradient boosting decision tree. *Energies* 12, 2522. doi:10.3390/en12132522
- Zhang, Z. L., Yang, Y., Zhao, H., and Xiao, R. (2022). Prediction method of line loss rate in low-voltage distribution network based on multi-dimensional information matrix and dimensional attention mechanism-long and short-term time-series network. *Transm. Distribution* 16, 4187–4203. doi:10.1049/gtd2.12590
- Zhou, S., Xue, J., Feng, Z., Dong, S., and Qu, J. (2022). “Abnormal line loss data detection and correction method,” in Proceedings of the 2022 4th asia energy and electrical engineering symposium (AEEES), Chengdu, China 832–837. doi:10.1109/AEEES54426.2022.9759815
- Zhu, L. P., and Lin, J. J. (2021). Learning spatiotemporal correlations for missing noisy PMU data correction in Smart grid. *IEEE Internet Things J.* 8, 7589–7599. doi:10.1109/jiot.2020.3040195



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Longchao Yao,
Zhejiang University, China
Wei Zhang,
Civil Aviation University of China, China

*CORRESPONDENCE

Guo Zhijun,
✉ guozhijun5270000@126.com

RECEIVED 28 November 2023

ACCEPTED 18 April 2024

PUBLISHED 15 May 2024

CITATION

Zhijun G, Weiming L, Qiuji C and Hongbo Z (2024), Terminal strip detection and recognition based on improved YOLOv7-tiny and MAH-CRNN+CTC models. *Front. Energy Res.* 12:1345574. doi: 10.3389/fenrg.2024.1345574

COPYRIGHT

© 2024 Zhijun, Weiming, Qiuji and Hongbo. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Terminal strip detection and recognition based on improved YOLOv7-tiny and MAH-CRNN+CTC models

Guo Zhijun^{1*}, Luo Weiming¹, Chen Qiuji¹ and Zou Hongbo²

¹Dongguan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Dongguan, Guangdong, China,

²College of Electric Engineering and Renewable Energy, China Three Gorges University, Yichang, China

For substation secondary circuit terminal strip wiring, low efficiency, less easy fault detection and inspection, and a variety of other issues, this study proposes a text detection and identification model based on improved YOLOv7-tiny and MAH-CRNN+CTC terminal lines. First, the YOLOv7-tiny target detection model is improved by the introduction of the spatially invariant multi-attention mechanism (SimAM) and the weighted bidirectional feature pyramid network (BiFPN). This also improves the feature enhancements and feature fusion ability of the model, balances various scales of characteristic information, and increases the positioning accuracy of the text test box. Then, a multi-head attention hybrid (MAH) mechanism is implemented to optimize the convolutional recurrent neural network with connectionist temporal classification (CRNN+CTC) so that the model could learn data features with larger weights and increase the recognition accuracy of the model. The findings indicate that the enhanced YOLOv7-tiny model achieves 97.39%, 98.62%, and 95.07% of precision, recall, and mean average precision (mAP), respectively, on the detection dataset. The improved MAH-CRNN+CTC model achieves 91.2% character recognition accuracy in the recognition dataset.

KEYWORDS

terminal strip, improved YOLOv7-tiny model, convolutional recurrent neural network with connectionist temporal classification, spatially invariant multi-attention mechanism, weighted bidirectional feature pyramid network, multi-head hybrid attention mechanism

1 Introduction

A more significant piece of insulating equipment (Huang et al., 2023) in the secondary equipment (Zhong et al., 2023) of a substation is the secondary circuit terminal strip. It serves as a line transmission component, connects the equipment inside and outside the screen, and carries numerous groups of mutually insulated terminal components. The ability of the protection device to connect to the main equipment via the terminal strip is crucial, and the ability of the protection device to operate normally is directly correlated with proper wiring. Normalizing the terminal block can significantly lower the likelihood of accidents resulting from the secondary circuit and the frequency of wiring errors. Current worker point-to-point inspections are not only slow but also prone to incorrect and inadequate inspections (Liu et al., 2023). The rapid advancement of deep learning (Wang et al., 2018) has led to a surge in the use of image detection and recognition in power-related fields, including live detection, robot inspection of substations, and unmanned aerial vehicle

inspection of transmission lines. These applications benefit from the high stability and accuracy of the recognition features of the technology.

Currently, the advancement of deep learning in the field of artificial intelligence technology has progressively established the mainstream. In order to avoid the hidden hazard of substation operation, Zhou et al. (2018) integrated the efficient and accurate scene text (EAST) algorithm into the line end identification of the screen cabinet to identify the text information. This algorithm was then combined with manual experience judgment. By employing combined character placement and recognition, Wang et al. (2020) increased the text character recognition accuracy and expedited the recognition process. The accuracy of each module cannot be optimized by this training strategy; it can only improve the model overall performance. Wang and Yi (2019) trained YOLOv5 by incorporating the concept of structural clipping into the model and then pruned the model based on the training outcomes. Training precision was lost even if the model scale was shrunk and training speed increased. The maximum pooling layer was applied to the convolutional neural network (CNN) model by Masci et al. (2012) in order to increase the model recognition accuracy. Cui et al. (2013) used a pattern of template matching to ascertain the direction and location of the terminal row; nevertheless, it is limited to determining these two factors, and manual labor is still required for text recognition. Wang et al. (2019) suggested fresh segmentation results, processed the algorithm, and employed a progressive way to segment texts of varying scales. Although the detection speed is lost, the detection rate is increased. Yang et al. (2022) described a text identification model that combines support vector machines with quad-pronged splitting having strong positioning effects and good accuracy, but it is too complex to extract design elements. Although the detection speed was poor, Xiaoxuan et al. (2021) built a set of intelligent recognition systems based on the YOLOv3 network and paired it with deep transfer learning approach. The multi-dimensional long short-term memory recurrent neural network with connectionist temporal classification (MDLSTM-RNN+CTC) model was proposed by Messina and Louradour (2015) and applied to text line character identification. This approach incorporated feature information from four dimensions thoroughly; however, its recognition accuracy was not very excellent.

In response to the shortcomings of the existing detection and recognition models, this paper presents an innovative approach for terminal text detection and recognition that combines an enhanced YOLOv7-tiny model with a MAH-CRNN+CTC architecture. Initially, the proposed improved YOLOv7-tiny object detection model integrates the spatially invariant multi-attention mechanism (SimAM), which plays a pivotal role in enhancing the model capacity to discern and focus on essential features of the target while filtering out noise, thereby boosting the overall detection performance. Subsequently, the model adopts a weighted bidirectional feature pyramid network (BiFPN), which efficiently consolidates feature maps from varying scales. This strategy enables the bidirectional exchange of feature information and dynamic allocation of weights according to feature significance, further refining the model precision in detecting targets. The MAH-CRNN+CTC recognition model introduces a multi-head attention hybrid (MAH) mechanism. This component facilitates

the comprehensive consideration of the entire sequence information, effectively addressing the issue of long-range dependencies. As a result, it accelerates the model training process, enhances feature extraction efficiency, and significantly boosts the model recognition accuracy.

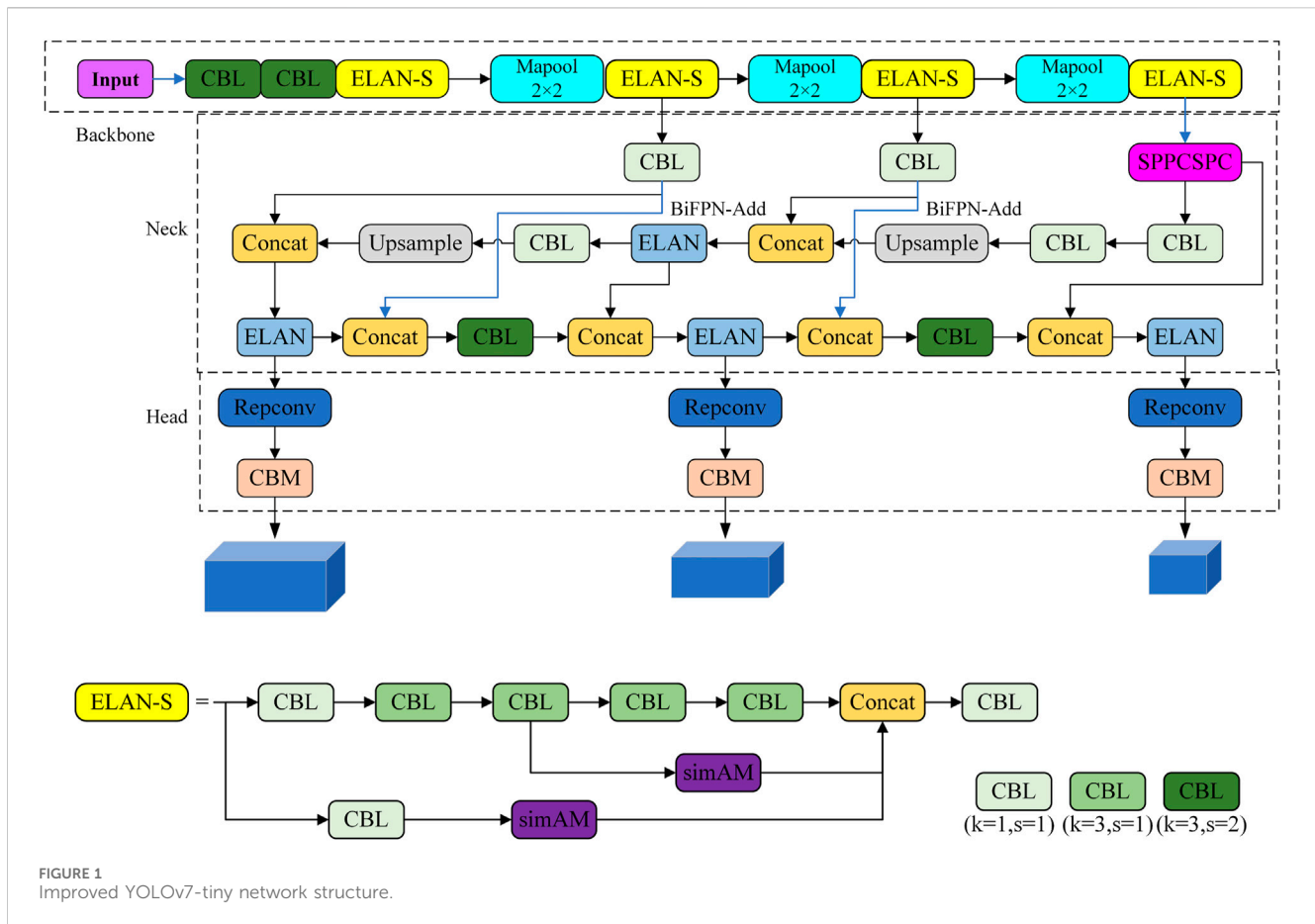
The bidirectional long and short-term memory (Bi-LSTM) module offers a potent temporal modeling tool that empowers the model to decipher and leverage intricate contextual cues within the input sequences, thereby bolstering both the precision and resilience of the recognition process. Conversely, the CNN module concentrates on achieving end-to-end text recognition through multi-level analysis and abstraction of images, transforming intricate image data into sequential features compatible with subsequent processing by the bidirectional Bi-LSTM component.

This paper is structured as follows: Section 1 introduces the text detection module, elaborating on the YOLOv7-tiny network model and detailing the improvements made to it; Section 2 encompasses an introduction to the text recognition module, focusing on the enhanced methods employed for improving the text recognition model; Section 3 presents the analysis of the experimental outcomes for both the text detection and recognition processes; and lastly, Section 4 presents a summary of the paper.

2 Text detection

2.1 Improving the YOLOv7-tiny network model

YOLOv7 consists of three components: the neck, which fuses features, the head, which makes predictions, and the backbone, which extracts features (Wu et al., 2019). The major components of the feature extraction network of the YOLOv7 network are the MPCnv, spatial pyramid pooling, cross-stage partial channel (SPPCSPC), E-ELAN, and Columbia Broadcasting System (CBS) modules. The E-ELAN module uses expand, shuffle, and merge cardinality to improve network learning while maintaining the original gradient path based on the original ELAN. After convolution of the feature map 3 times and 5×5 , 9×9 , and 13×13 maximum pooling, the SPPCSPC module uses the concept of spatial pyramid pooling to obtain image features under various receptive fields. This solves the issue of repetitive feature extraction from the image by the convolutional neural network. Subsequently, the characteristics of distinct receiving domains are combined, and following double convolution, they are ultimately split with the feature map. The MPCnv module uses a 2×2 maximum pooling operation to increase the receptive field of the current feature layer. It then uses 1×1 convolution to adjust the number of channels. Finally, it fuses the feature information that has been processed with the feature information obtained by normal convolution to improve the feature extraction capability of the network. As the YOLOv7 feature fusion network, the path aggregation network (PANet) is utilized to fuse the deep semantic and shallow location characteristics of the image and produce feature maps of various sizes. The RepConv structure modifies the number of channels for characteristics with varying scales on the prediction side.



YOLOv7-tiny is an improvement over YOLOv7, as shown in Figure 1. ELAN-S is utilized in place of E-ELAN in the backbone section, and the SimAM module is added to the ELAN-S structure to improve the feature expression capabilities of the network. The Max pooling operation is exclusively used for down sampling, and the convolution process in MPConv is canceled. The BiFPN module is incorporated into the SPPCSPC structure for feature fusion in the neck section. The RepConv structure is still used in the head section to modify the number of channels for features with varying scales.

2.2 SimAM module

The module for attention mechanisms different from the channel attention mechanism and spatial attention that have been previously proposed, SimAM (also known as the SimAM module) (Yang et al., 2021) is a lightweight attention module that is both simple and incredibly effective. The SimAM module will not add further complexity to the network because it does not include any extra parameters. It is a feature map-derived 3D attention method. This module uses the energy function to optimize it in accordance with neuroscience theory and quickly arrive at an analytical solution; in other words, it uses the energy function to determine the attention mechanism weight. The energy function $e_t(\cdot)$ is defined as follows:

$$e_t(w_t, b_t, y, x_i) = (y_i - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2, \quad (1)$$

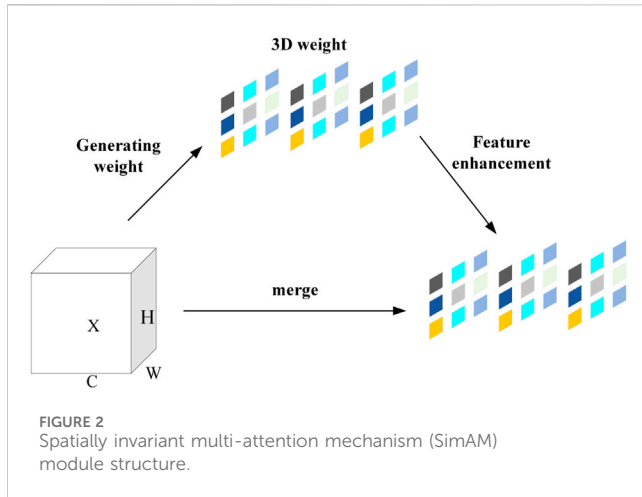
$$\begin{cases} \hat{t} = w_t t + b_t \\ \hat{x}_i = w_t x_i + b_t \end{cases}, \quad (2)$$

where t and x_i are the target neuron and other neurons of the input feature tensor X , respectively, and $X \in \mathbb{R}^{C \times H \times W}$. C , H , and W are the pass number, height, and width of the feature tensor, respectively. i is the neuron index on a certain number of channels. M is the number of neurons on the channel, $M = H \times W$. w_t and b_t target neuron transform weights and bias, respectively. y , y_t , and y_0 are scalar quantities, of which y_t and y_0 are for different values; this paper introduced the binary label instead, with $y_t = 1$ and $y_0 = -1$.

Neurons inside the same channel can be trained to have their linear separability minimized by minimizing Eq. 1. By incorporating a regular term and employing binary labels, the energy function can be changed to

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} [-1 - (w_t x_i + b_t)]^2 + [1 - (w_t t + b_t)]^2 + [1 - (w_t t + b_t)]^2 + \lambda w_t^2, \quad (3)$$

where λ is the regularization coefficient and w_t is the weight of the transformation of the i neuron.



Eq. 3 to Eq. 4.

$$\begin{cases} w_t = -\frac{2(t - u_t)}{(t - u_t)^2 + 2\sigma_t^2 + 2\lambda} \\ b_t = -\frac{1}{2}(t + u_t)w_t \end{cases}, \quad (4)$$

where u_t and σ_t^2 are both intermediate variables. Eq. 5,

$$\begin{cases} u_t = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i \\ \sigma_t^2 = \frac{1}{M-1} (x_i - u_t)^2 \end{cases} \quad (5)$$

By substituting w_t and b_t into Eq. 2, the minimum energy e_t^* can be obtained, that is Eq. 6,

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda}. \quad (6)$$

u_t and σ_t^2 are replaced with the mean $\hat{u} = \frac{1}{M} \sum_{i=1}^M x_i$ and variance $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{u})^2$, respectively.

The lower the energy, the greater and more important the difference between the target neuron and the peripheral neuron t . The importance of neurons can be obtained by obtaining $\frac{1}{e_t^*}$, and then the enhanced feature tensor \tilde{X} can be obtained by using Eq. 7:

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X, \quad (7)$$

where X is the input characteristic tensor. E is the sum of e_t^* in all channels and spatial dimensions. \odot is the Hadamard product.

In Eq. 7, the sigmoid function is added to limit the excessive value of E , and the sigmoid function does not affect the relative importance of each neuron.

Figure 2 shows the SimAM module chart. It can be seen as a cell aimed at increasing the convolution characteristic expression ability of the neural network; any intermediate feature tensor can be taken as the input and the transformation output with the same size and have the feature of enhancing the characterization of the tensor, where X is the input feature tensor in the figure.

The biggest advantage of this module is based on the defined energy function to choose from.

2.3 Weighted bidirectional feature pyramid network

As shown in Figure 3A top-down pathway of the feature pyramid network (FPN) allows for feature fusion. A certain amount of detection accuracy can be increased by the fused high-level semantic information. Prior feature fusion techniques frequently treated the feature information of various scales identically. Although it is impossible to determine the relative relevance of many input features, each contributes differently to the output features. This implies that the characteristics of some scales might be more significant and have a bigger influence on the outcome. Consequently, the weighted BiFPN is proposed in this research (Tan et al., 2020). As shown in Figure 3B, additional weights are applied for each input, utilizing a distinct blend of several input properties.

First, the nodes with a single input edge and little contribution are eliminated to simplify the network and decrease the amount of parameters. This effectively lowers the network complexity. Second, based on the properties of three distinct scales, the jump connection mechanism was established, increasing a feature fusion path in the quantity under the assumption of somewhat larger. Diagrams will be used to better integrate low-level and high-level semantic information, and weights can be used to focus network model studies on the most important informational properties, thereby enhancing network performance and characterization. The calculation of wighted feature fusion in BiFPN is represented by Eq. 8:

$$\text{Out} = \sum_{i=0} \frac{\omega_i * I_i}{\varepsilon + \sum_{j=0} \omega_j}, \quad (8)$$

where ω represents the learnable weight, I_i represents the input feature, and $\varepsilon = 0.0001$.

3 Text recognition methods

3.1 Improved CRNN+CTC algorithm

Text recognition is all that is needed to identify secondary device terminals. The convolutional recurrent neural network (CRNN) model not only performs well for more complicated texts, handwritten letters, and symbols but it also does not require segmenting the target to precisely mark the characters. It also has no restrictions on the length of the text sequence. There are not many model parameters, and training proceeds quickly. The model structure is thereby enhanced and optimized by making a reference to the network architecture of the traditional text recognition model or CRNN. Meanwhile, to better mine the long-distance data features of correlated time series, the MAH mechanism is introduced to the Bi-LSTM in the recurrent neural network module to accommodate the secondary equipment terminal strip identification of the substation. The network architecture of the substation secondary

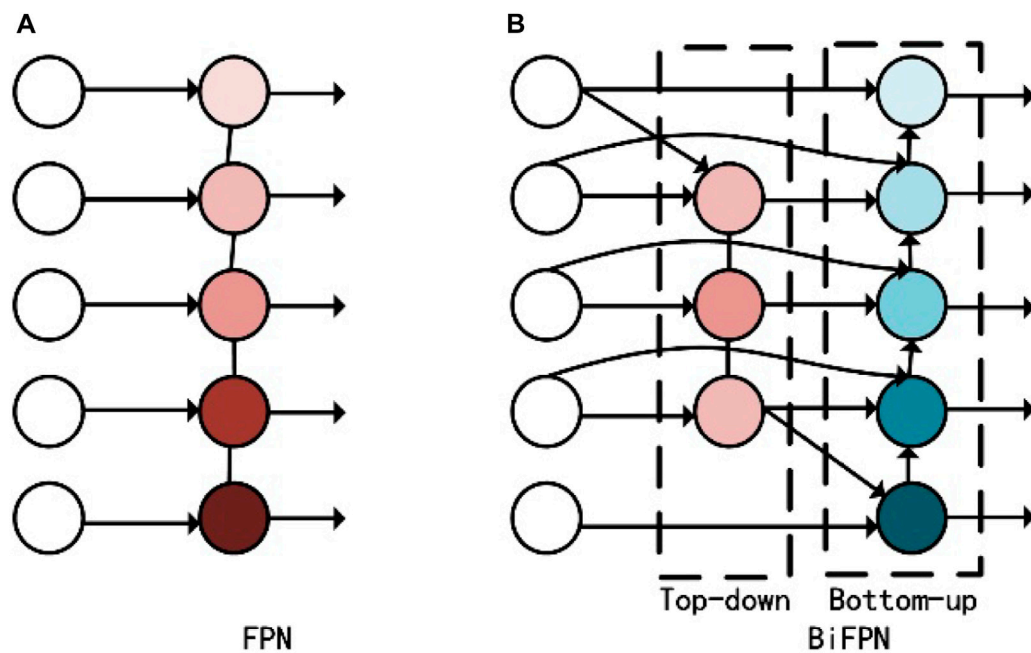


FIGURE 3
Improved feature structure pyramid.

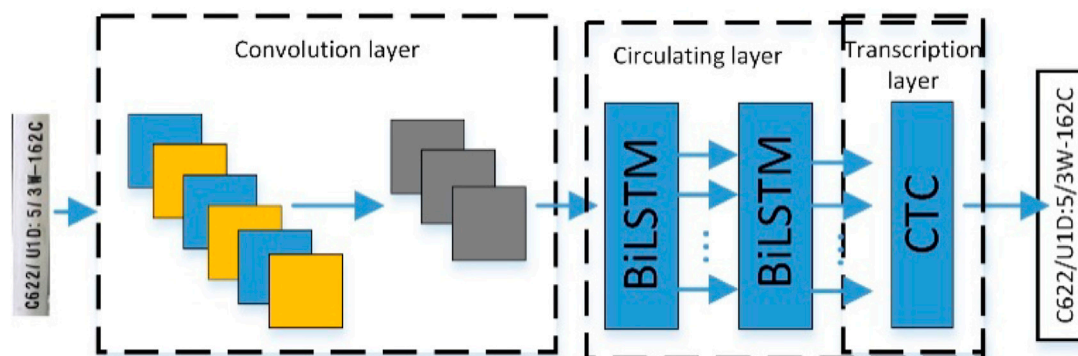


FIGURE 4
Network architecture of the identification model of the secondary equipment terminal strip in the substation.

equipment terminal strip recognition model is shown in Figure 4. The three primary components of the terminal strip recognition model are the connection temporal classification (CTC), Bi-LSTM neural network, and convolutional neural network (CNN). These include the CNN for picture feature extraction, the Bi-LSTM for character sequence extraction, and CTC for character mismatch resolution.

3.2 Feature extraction network

The third and fourth max pooling kernel scales in the CNN module are set to 1×2 pixels, making it simple to use the CNN features that have been extracted as the recurrent neural network (RNN) input. To

expedite the network training process, batch normalization layers are incorporated after the fifth and sixth layers of convolution. The original image height will be decreased to a fixed value of 32 pixels before it is entered into the CNN. The width of each feature vector in the feature sequence is set to 1 pixel, and they are all generated in the same direction as the feature map sequence, that is, from left to right. The first feature vector is linked to the first feature map.

In order to increase the network training speed, the BN layer is added to the CNN module in this research. The variable body of the ReLU function, known as the Leaky ReLU function, is adopted by the activation function. To address the issue of neurons not learning after the negative interval of the ReLU function, a leak value is added to the negative interval of the ReLU function, causing the output to slope slightly toward the negative input. As shown in Table 1, the CNN

TABLE 1 CNN network structure.

Network layer	Input size
Convolution layer	$64 \times 32 \times 160$
Maximum pooling layer	$64 \times 16 \times 80$
Convolution layer	$128 \times 16 \times 80$
Maximum pooling layer	$128 \times 8 \times 40$
Convolution layer	$256 \times 8 \times 40$
Maximum pooling layer	$256 \times 4 \times 40$
Convolution layer	$512 \times 4 \times 40$
Maximum pooling layer	$512 \times 2 \times 40$
Convolution layer	$512 \times 1 \times 40$

module gains 4 maximum pooling layers in this study, with the final 2 pooling layers having convolution kernel sizes of 1×2 pixels. The remaining convolution kernel sizes are 3×3 pixels and padding = 1, with the exception of the final convolution layer, which has a convolution kernel size of 2×2 pixels and padding = 0. The input image is processed in this article to create a 32×160 -pixel image. After the CNN, the resulting feature map size is $512 \times 1 \times 40$ pixels, meaning that there are 512 feature maps in total, each with a height of 1 pixel and a width of 40 pixels.

3.3 Sequence prediction network

The sequence properties in the sequence label distribution of each frame are predicted using the model prediction module. The RNN is highly proficient at capturing contextual connections in the realm of sequential text recognition. However, while processing lengthy texts, the standard RNN loses its ability to connect distant information and becomes vulnerable to the gradient disappearance issue, which makes the network difficult to converge and results in

low training accuracy. An exceptional variety of the RNN that excels at acquiring long-term dependent data is the long short-term memory (LSTM). It can selectively recall the information that must be retained for a long time and forget the irrelevant information. It can also regulate the information transferred through the gate empty state, as shown in Figure 5.

Through its forgetting gate, input gate, output gate, and other gating structures, the LSTM cell structure may efficiently save and regulate the cell state update; the update rules are shown in Eq. 9. Eqs 9–14 illustrate how the gating unit is realized.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (9)$$

$$f_t = \sigma(W_f[H_{t-1}, x_t] + b_f), \quad (10)$$

$$i_t = \sigma(W_i[H_{t-1}, x_t] + b_i), \quad (11)$$

$$\tilde{C}_t = \tanh(W_c[H_{t-1}, x_t] + b_c), \quad (12)$$

$$O_t = \sigma(W_o[H_{t-1}, x_t] + b_o), \quad (13)$$

$$H_t = O_t * \tanh(C_t), \quad (14)$$

where H_{t-1} is the output of the hidden layer of the previous unit. x_t is the input of the current cell. f_t , i_t , and O_t represent the output of structures such as forgetting, input, and output in the gating structure, respectively. C_t , C_{t-1} , and \tilde{C}_t represent the cell state of the current moment, the cell state of the previous moment, and the cell state of the output layer, respectively. W_f , W_c , and W_o are the corresponding weight parameters of the gate, respectively. b_f , b_i , b_c , and b_o are the bias parameters corresponding to the gate, respectively. $[\cdot]$ is the vector connection symbol.

While the feature sequence recognition of the secondary device terminal number considers both the past and future context information to be beneficial, one-way LSTM only employs the past context information. Consequently, this article employs the Bi-LSTM network module, which uses the future information backward and the past information forward, as shown in Figure 6.

In this paper, a two-layer Bi-LSTM is set up. The output of the CNN is a feature map of size $m \times T$, where T is the output sequence length of the feature module and m is the number of channels. After

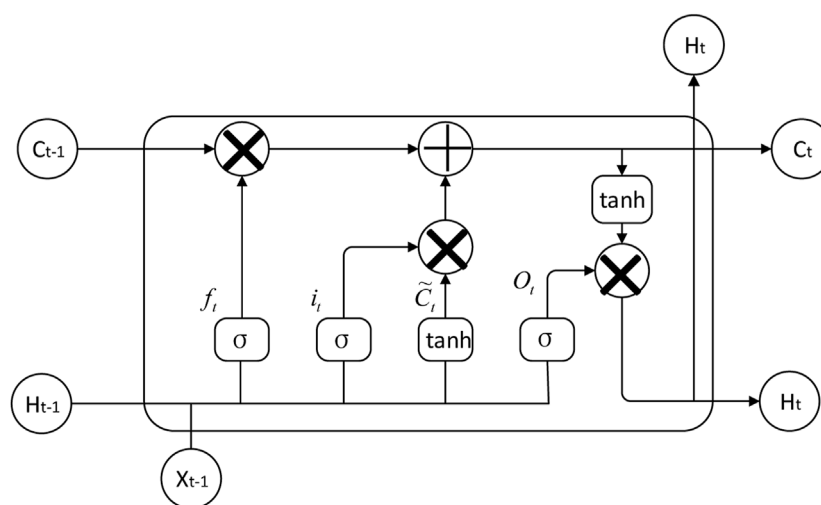
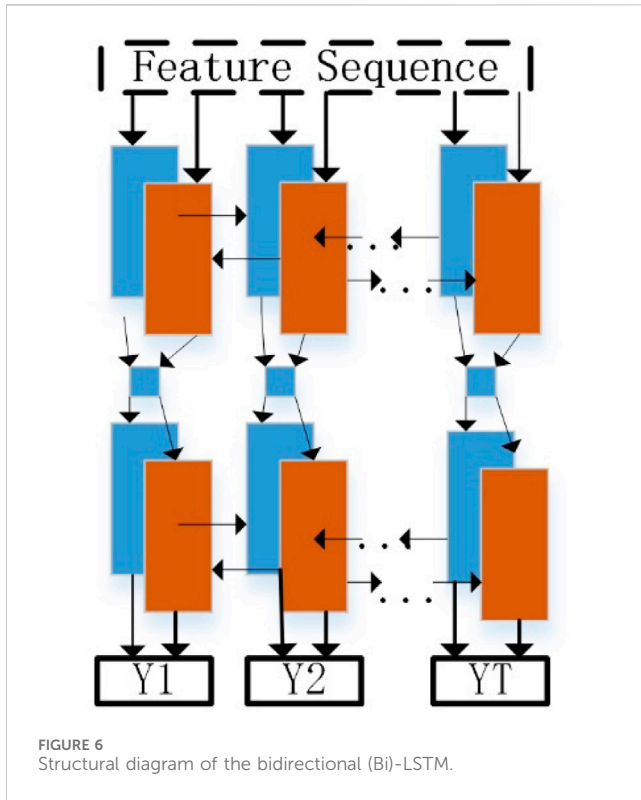


FIGURE 5
Long short-term memory (LSTM) cell structure.



transforming $x = (x_1, x_2, \dots, x_T)$ of each column through “map to sequence,” it is input into the Bi-LSTM module and the output vector $T \times n_{class}$ of length $y = (y_1, y_2, \dots, y_T)$, where n_{class} is the number of sub-row character categories of the secondary device terminal.

3.4 Multi-head attention hybrid mechanism

This article presents a model in a lengthy attention mechanism that supplements the Bi-LSTM module. This helps the Bi-LSTM module better address the correlation characteristic of long time-series data mining as the problem of long sequences making it easier to lose information arises during the training process. The output vector is transformed into three input matrices of dimension d_k by three different mapping operations, Q (Query), K (Key), and V (Value), and the attention output matrix is given by Eq. 15:

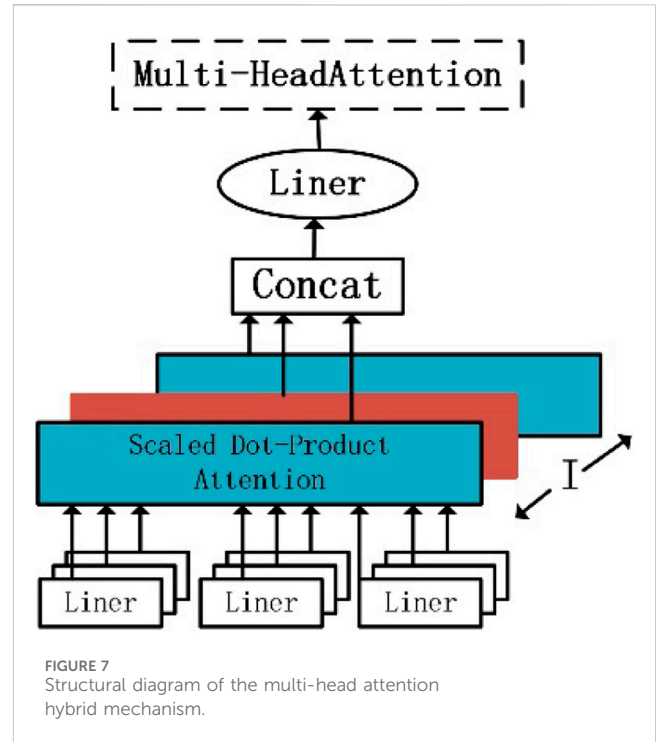
$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (15)$$

where d_k is the feature dimension of each key, which is used for weight scaling, and softmax is normalized to the interval [0,1].

The multi-head attention hybrid mechanism divides the time series into an I subspace, and each head performs self-attention calculation on the subspace to enhance its expressive power. Then, the results of head I are spliced and integrated to obtain multiple heads, and each head is splice to obtain the final through linear transformation, that is Eq. 16, 17.

$$heads_i = Attention(QW_i^Q, KW_i^K, VW_i^V), \quad (16)$$

where W_i^Q , W_i^K , and W_i^V represent the weight matrix of Q, K, and V, respectively.



$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_I)W^o, \quad (17)$$

where W^o represents the weight of the linear transformation; $head_i$ represents head i in the bull attention module; and $Concat$ represents the splicing operation. $MultiHead(Q, K, V)$ is the final output result, which can learn more feature information from different spaces, and its model structure is shown in Figure 7:

3.5 Transcription layer

The problem of difficult-to-align input and output is a common occurrence in the text recognition sector. Thus, in this article, the recurrent neural network is decoded using CTC, and the Bi-LSTM output is transformed into a sequence format.

π is defined as the text sequence path composed of the Bi-LSTM output. For the Bi-LSTM module, the probability of output x given input l is calculated by the following Eq. 18:

$$p(l|x) = \sum_{\pi \in \beta^{-1}(l)} p(\pi|x), \quad (18)$$

where β is a multi-to-one mapping function, the purpose of which is to remove duplicate labels and blank labels. $\pi \in \beta^{-1}(l)$ represents all l paths that are π after transformation, and any path, as shown in Eq. 19.

$$p(\pi|x) = \prod_{t=1}^T y_{\pi_t}^t, \forall \pi \in L^T, \quad (19)$$

where T represents the length of the input sequence and l is the label of the output. π_t represents the output character corresponding to path π at time t , which corresponds to the probability of obtaining the character at time t .

TABLE 2 Comparison of experimental results for text detection.

Model	P/%	R/%	Mean average precision (mAP)@0.5/%	Model size/MB	FPS (f/s)
YOLOXs (Yin et al., 2023)	95.58	79.14	87.21	16.4	86.90
YOLOv4-tiny (Zhao et al., 2023)	83.57	73.06	74.30	22.5	77.41
YOLOv5s (Han et al., 2024)	91.67	79.35	85.66	14.5	83.95
YOLOv7-tiny	94.91	84.82	92.15	12.2	103.42
Improved YOLOv7-tiny	97.39	89.62	95.07	12.08	95.87

The bold values represents the improved experimental results of this paper.

TABLE 3 Comparison of methods.

Model	P/%	R/%	mAP@0.5/%
YOLOv7-tiny	94.91	84.82	92.15
YOLOv7-tiny + BiFPN	94.89	87.42	93.16
YOLOv7-tiny + SimAM	96.16	87.01	93.36
YOLOv7-tiny + BiFPN + SimAM	97.39	89.62	95.07

The bold values represents the improved experimental results of this paper.

In Eq. 20, the training process of CTC is to adjust the parameter $\frac{\partial p(l|x)}{\partial \omega}$ of Bi-LSTM through the gradient ω so that $\pi \in \beta^{-1}(l)$ is maximized when the input sample $p(l|x)$ is obtained as

$$h(x) = \arg \max_{l \in L \leq T} p(l|x).$$

(20)

4 Analysis of experimental results

4.1 Experimental environment

In this paper, the operating system used for model training is Windows 10 with a 64-bit processor. Intel(R) Core(TM) i5-10200H CPU @ 2.40 GHz 2.40 GHz is used as hardware. Running memory was 12 GB. PyTorch is chosen as the deep learning framework. The programming language is Python 3.6. The CUDA version is 11.6.

4.2 Evaluation index

In this paper, precision (P), recall (R), and mean average precision (mAP) are used as evaluation indicators for text detection, as shown in Eqs 21–24.

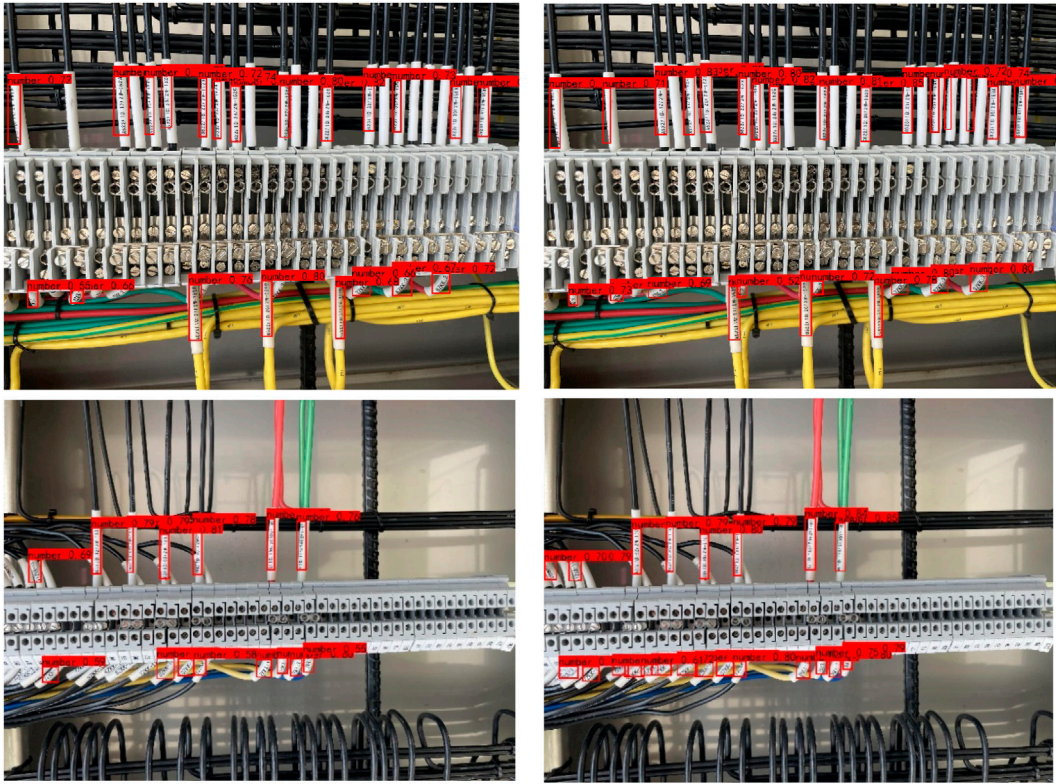


FIGURE 8 Comparison of the terminal strip text detection effect.

TABLE 4 Comparison with other methods.

Model	ACC (%)	Loss
CRNN+CTC	85.4	0.0981
MAH-CRNN+CTC (Guo et al., 2022)	87.59	—
Improved MAH-CRNN+CTC	91.2	0.0329

The bold values represents the improved experimental results of this paper.

$$P = \frac{TP}{TP + FP} \times 100\%, \quad (21)$$

$$R = \frac{TP}{TP + FN} \times 100\%, \quad (22)$$

$$AP = \int_0^1 P(R)dr, \quad (23)$$

$$mAP = \frac{\sum_{j=1}^S AP_j}{S}, \quad (24)$$

where TP stands for true positive, indicating the samples that the network detects following detection and classification match samples that have been labeled. False negatives or labeled samples that the network did not detect or classify—also known as missed detection—are represented by FN and FP , respectively. False positives are incorrectly classified detection samples that are not included in the labeled samples or false detection. The average precision (AP) of a single class is the area measured between the $P(R)$ curve and the axis. Averaging the AP s of all categories yields the mAP , where S is the total number of categories.

Average loss (Loss) and character recognition accuracy (Acc) are often used evaluation metrics for text recognition. Eq. 25, which illustrates the condition of incorrect recognition and multiple recognitions, shows that Acc is the ratio of the number of characters identified by model A to the total number of characters identified by model B. The average loss of character recognition is shown in loss. The better the model, the larger the Acc, and the smaller the loss.

$$A_{cc} = \frac{A}{B}. \quad (25)$$

4.3 Text detection experiment and result analysis

The dataset employed in the text detection module within this paper is sourced from a collection of 1,000 high-resolution images

(1,024 × 1,024), depicting terminal rows of secondary equipment in substations. The division of data in this set follows a 8:2 ratio for training and testing subsets, respectively; moreover, the training subset itself is further stratified into a training set and a validation set according to a 9:1 allocation principle.



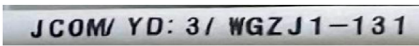
In the experiment, the epoch is set to 200, batch size is 8, the Adam optimizer is used to update the optimization gradient, and the cosine annealing algorithm is used to dynamically adjust the learning rate attenuation strategy. The initial learning rate of the model is 0.001, the weight attenuation parameter is 0.0005, and the learning rate momentum parameter is 0.937.

To confirm that the revised model presented in this work is superior, Table 2 compares the revised model with lightweight models like YOLOXs (Yin et al., 2023), YOLOv4-tiny (Zhao et al., 2023), YOLOv5s (Han et al., 2024), and YOLOv7-tiny based on the terminal strip wiring dataset of secondary devices. The enhanced model in this study has an average accuracy (mAP) of 95.07%, which is 7.86%, 20.77%, 9.41%, and 2.92% greater than that of YOLOXs, YOLOv4-tiny, YOLOv5s, and YOLOv7-tiny, respectively, based on the experimental findings shown in Table 2. With a memory occupation of only 12.08 MB, the upgraded model outperforms the YOLOv4-tiny model by 46.3%. In order to guarantee accuracy, the enhanced model outperforms the other models in terms of accuracy and recall rate, that is, by 97.39% and 89.62%, respectively. The average detection speed (FPS) of the enhanced model is 95.87 f/s, which is marginally slower than the quickest detection speed of YOLOv7-tiny; nevertheless, this model performs better in other detection algorithm performance tests. Therefore, the upgraded model in this research still exhibits significant improvements in the identification of speed and accuracy with respect to the total detection performance of the model.

An array of ablation experiments was created for comparison analysis in order to confirm the efficacy of the modified YOLOv7-tiny model suggested in this paper. The trials were carried out using the same training conditions to guarantee the accuracy of the experiments. The comparative findings are shown in Table 3 for the original model, each upgraded module, and the test set.

Table 3 shows how the precision rate, recall rate, and mAP increased by 1.25%, 2.91%, and 1.21%, respectively, when the SimAM was added to the original model. It demonstrates that compared to the original model, the SimAM module is more capable of feature extraction and expression. The precision, recall, and mAP of the model improved to 94.89%, 87.42%, and 93.16%, respectively, after the FPN module was swapped out for the BiFPN module in the neck network. This improvement was

TABLE 5 Comparison of recognition renderings of the model.

Picture	CRNN+CTC	Improved MAH-CRNN+CTC
	2-32KK1-0	2-32KK1-6
	J04/YD:13/WGZJ1-181	J04/YD:13/WGZJ1-131
	JCOM/YD:3/WGZJ1-131	JCOM/YD:3/WGZJ1-131

The bold values represents the improved experimental results of this paper.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

GZ: writing–review and editing. LW: writing–review and editing. CQ: writing–review and editing. ZH: writing–original draft.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China under Grant No. 52107108.

References

- Cui, Y. X., Li, Y., Wang, H. J., and Wang, X. L. (2013). Identification of location and orientation for terminal blocks based on template matching. *Key Eng. Mater. Zurich* 561, 515–520. doi:10.4028/www.scientific.net/kem.561.515
- Guo, Ke, Bai, Y., Shao, X., Wang, X., and Ma, J. (2022). Terminal block text recognition based on multi-scale attention and convolutional recurrent neural networks. *J. Anhui Univ. Nat. Sci. Ed.* 46 (06), 49–56.
- Han, C., Yang, ZHOU, Wang, L., Lei, H., Yao, D., and Liang, W. (2024). Research of CCTV drainage pipeline defect recognition method based on YOLOv5s. *Munic. Technol.* 42 (0 3), 230–236. doi:10.19922/j.1009-7767.2024.03.230
- Huang, H., Wu, J., Xiao, H., Liang, Z., Wang, J., Tan, X., et al. (2023). Terminal text detection and recognition based on attention mechanism. *Mech. Electr. Eng.* 52 (06), 202–206.
- Liu, W., Lin, G., Fu, D., and Wang, S. (2023). Substation secondary loop terminal row design text detection and recognition. *J. hubei Univ. Natl. Nat. Sci. Ed.* 9 (02), 198–205. doi:10.13501/j.carol.carroll.nki.42-1908/n.2023.06.009.06
- Masci, J., Meier, U., Ciresan, D., Schmidhuber, J., and Fricout, G. (2012). Steel defect classification with max-pooling convolutional neural networks. *2012 Int. Jt. Conf. Neural Netw. IJCNN*, 1–6. doi:10.1109/ijcnn.2012.6252468
- Messina, R., and Louradour, J. (2015). “Segmentation-free hand-written Chinese text recognition with LSTM-RNN,” in Proceeding of the 13th IAPR International Conference on Document Analysis and Recognition, Nancy, France, August, 2015, 171–175.
- Tan, M., Pang, R., and Le, Q. V. (2020). “Efficient det: scalable and efficient object detection,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, June, 2020, 10781–10790.
- Wang, X., and Yi, Z. (2019). Research on Obstacle detection method of Mowing robot Working environment based on improved YOLOv5. *Chin. J. Agric. Mech.* 44 (3), 171–176.
- Wang, J., Ma, Y., Zhang, L., Gao, R. X., and Wu, D. (2018). Deep learning for smartmanufacturing: methods and applications. *J. Manuf. Syst.* 48, 144–156. doi:10.1016/j.jmsy.2018.01.003
- Wang, L., Huang, Li, Zhang, L., Long, Z., Li, Y., and Zhou, J. (2020). Fault diagnosis technology of transformer substation panel cabinet based on joint training method. *Electr. Power Supply* 37 (Suppl. 5), 85–90.
- Wang, W., Xie, E., Li, X., Hou, W., Lu, T., Yu, G., et al. (2019). “Shape robust text detection with progressive scale expansion network,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, June, 2019, 9336–9345.
- Wu, X., He, Y., Zhou, H., Cheng, L., and Ding, M. (2019). Research on environmental personnel identification of monitored waters based on improved YOLOv7 algorithm. *J. Electron. Meas. Instrum.* 37 (5), 20–27.
- Xiaoxuan, Hu, Xijin, Z., Qi, Z., and Wang, H. (2021). Intelligent Detection system of Marine Welding Surface defects based on deep transfer learning. *Shipbuild. Technol.* 49.
- Yang, Z., Huang, H., He, L., Liu, Z., Li, X., and Liu, Z. (2022). Surface defect detection of circuit board based on color histogram. *Comput. Integr. Manuf. Syst.*,
- Yang, L., Zhang, R. Y., Li, L., and Xie, X. (2021). “SimAM: a Simple, Parameter-free attention module for convolutional neural networks,” in Proceedings of the 38th International Conference on Machine Learning, Virtual, July, 2021, 11863–11874.
- Yin, Z., Qi, Y., and Wang, L. (2023). TERMINAL_PIN welding surface defect detection method based on YOLOXs. *Comput. Appl.* 43 (S2), 209–215.
- Zhao, J. G., Han, Z. S., Fan, J. J., and Zhang, J. (2023). Safety helmet wearing detection algorithm based on improved YOLOv7-tiny. *J. Hebei Univ. Archit. Eng.* 41 (04), 240–245.
- Zhong, M., Jun, T., Fu, A. M., and Yang, Y. (2023). Based on attention mechanism of secondary loop terminal text detection and recognition method. *Electr. power Sci. Technol. J.* 38 (03), 132–139. doi:10.19781/j.issn.1673-9140.2023.03.014
- Zhou, J., Yan, Li, and Zhang, K. (2018). Intelligent identification System for line sleeve label of screen cabinet equipment in substation secondary system. *Mach. Electron.* 36 (11), 67–70.

Conflict of interest

Authors GZ, LW, and CQ were employed by Dongguan Power Supply Bureau of Guangdong Power Grid Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Nishant Kumar,
Indian Institute of Technology Jodhpur, India
Juan Wei,
Hunan University, China

*CORRESPONDENCE

Wenbiao Tao,
✉ 250254240@qq.com

RECEIVED 03 February 2024

ACCEPTED 18 April 2024

PUBLISHED 30 May 2024

CITATION

Meng T, Lu X, Wang X, Wang L, Tao W, Shan L
and Geng X (2024), A fusion topology method
for generating new equipment startup schemes
for power grids.
Front. Energy Res. 12:1381332.
doi: 10.3389/fenrg.2024.1381332

COPYRIGHT

© 2024 Meng, Lu, Wang, Wang, Tao, Shan and
Geng. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

A fusion topology method for generating new equipment startup schemes for power grids

Tao Meng¹, Xiaohui Lu¹, Xiaoang Wang¹, Liang Wang²,
Wenbiao Tao^{2*}, Lianfei Shan³ and Xiaofei Geng³

¹State Grid Shanxi Electric Power Company, Shanxi, China, ²State Grid Shanxi Electric Power Research Institute, Shanxi, China, ³Beijing Kedong Electric Power Control System, Beijing, China

New grid equipment startup programs are widely used in various countries to regulate the commissioning of new equipment; these programs have unique differences in terms of strictness, information asymmetry, and complexity relative to other types of startup programs. With respect to rule-based generation methods, because the method of revising the rules weakens their migration ability, it is difficult to adapt these methods to the *status quo* of high-speed power grid construction; moreover, most of the current generation methods based on deep learning improve upon the rule-based methods but do not eliminate the rules of the constraints. Therefore, this paper presents a fusion topology for generating a new grid equipment startup scheme, which generates the scheme from end to end. The method utilizes the powerful processing capabilities of the GATv2 model and the ERNIE-GEN model for topology and text, respectively. The device type-based coding strategy and the scheme complexity-based self-attention layer selection strategy are used in the GATv2-based device identification model to address information asymmetry and complexity variability, and the device information modification strategy is applied to solve the strictness variability problem in the ERNIE-GEN-based scheme generation model. Finally, through the testing and verification of field data from four types of new equipment startup schemes in real power grids, it is verified that the method can effectively generate new equipment startup schemes for power grids, and the reasonableness and efficiency of the three strategies are verified through ablation experiments, which verify that the method can effectively generate new equipment startup schemes for power grids that meet the requirements of real power grids.

KEYWORDS

GATv2, ERNIE-GEN, new equipment operation, start-up plan, deep learning

1 Introduction

The power system is an intricate network that covers power generation, transmission and distribution. As the nerve center of modern society, the power system has experienced unprecedented high-speed development in recent years. With the rapid progress of science and technology and the continuous growth of energy demand, as well as to meet the construction requirements of smart grids (Chen et al., 2009; Zhang et al., 2009; Dong et al., 2014; Yi et al., 2009; Yu and Luan, 2009), power grids need to not only have a larger capacity and higher efficiency but also need to constitute a more intelligent mode of operation, which leads to a large number of new equipment access requirements. However, this large-scale

equipment access also brings a series of challenges for the power grid; whenever a grid accesses new equipment, this may have an impact on the structure, working mode, and stability of the current power grid, and if there is a stability problem in the power grid when accessing new equipment, this may lead to power supply interruptions, equipment damage or even serious accidents, which may have an impact on the stability and operation of society in general (Sanjab et al., 2016; Otuoze et al., 2018; Sun et al., 2018). Therefore, the development of a new grid equipment startup program is crucial, and for actual new equipment access operators, the rationality of the new grid equipment startup program is directly related to personnel safety; thus, new grid equipment startup programs are widely used in various countries as a kind of grid operation ticket for regulation (Li et al., 2010; Mengchao et al., 2010; Harbor et al., 2014; Wang et al., 2014; Ren et al., 2022).

Grid operation tickets are formal documents or recording sheets used in power systems to perform, record, and control grid operations. These operation tickets are usually written guidance or recording tools used when performing various critical operations in power system operation (Zhou and Yang, 2004; Zhou et al., 2004; Liu et al., 2005; Yuan et al., 2022), and their purpose is to ensure that operators perform grid operations, maintenance, switching operations, etc., according to specified procedures and standards to maintain the safe, stable and reliable operation of the grid (Zhu et al., 2003; Wu et al., 2006; Zhu et al., 2022). Initially, the grid operation ticket is manually written by personnel. At this time, the correctness of the grid operation ticket depends on the experience of specific personnel, and the degree of uncertainty is great; however, with the development of science and technology and the continuous construction of power grids and intelligent methods, especially given the rapid development of computer technology, the generation of grid operation tickets has entered the stage of automatic generation. With several years of research on methods of generating grid operation tickets, these methods are divided into rule-based generation (Gong et al., 2006) and neural network-based deep learning generation (Cai and Qi, 2021).

The rule-based generation of grid operation tickets involves generating operation tickets using predefined rules and processes; this method relies more on the accurate specification of grid operations, including the steps, conditions, and limitations of various operations. By developing detailed operating procedures and standardized processes, the process of generating operating tickets can be made more controllable and deterministic. The current rule-based methods for grid operation ticket generation involve expert rules (Tang et al., 2001) and multiple intelligent methods (Zhou et al., 2004). Expert rules are sets of rules and guidelines for generating grid operation tickets based on the experience and knowledge of experts in the domain of power systems (Song, 1999). Hu et al. (Hu et al., 2002) simplified the expert system knowledge base by establishing a generalized cognitive model of power system equipment, which was applied in several substations. Yang et al. (Yang et al., 2004) established an operation based on objects and designed an expert system with a network topology, a knowledge base, and a reasoning mechanism suitable for power system dispatch operations. An intelligent system is usually represented as a computer program that interacts with the outside world through a preset protocol. In a multi-intelligent system,

through distributed decision-making, each intelligent body generates corresponding operation tickets according to the part it is responsible for, and they are coordinated in the multi-intelligent system to ensure the consistency and efficiency of the whole grid. Even if one intelligent body fails or is temporarily disabled for some reason, other intelligent bodies can temporarily take over the part that this intelligent body is responsible for to prevent the system from failing and to ensure that the operation tickets can be created. Preventing the system from failing ensures that the generation of operation tickets is not disturbed (Jiang et al., 2005; Li et al., 2016; Yonggang et al., 2016). Guo (Guo et al., 2006) and others built an operation ticket generation system based on the fuzzy cognitive map reasoning model of intelligence. Wei (Wei et al., 2023) and others proposed an optimal fault recovery control (TROFC) scheme for WF acquisition systems based on topology reconstruction. Liu (Liu et al., 2023) and others proposed a prediction method based on Adaboost ensemble convolutional neural network and bidirectional long short-term memory. In addition, there are many rule-based generation methods, such as that of An et al. (2021), who used association rule algorithms to mine historical ticket information and proposed a method to establish a knowledge base of historical tickets. Overall, this rule-based generation method achieves some effectiveness in the region where it is initially constructed, but its cumbersome rule revision process, as well as strong specialization, often necessitates changing the use of the rules when they are migrated elsewhere.

The generation method based on neural network deep learning generation utilizes the historical and online data of the power grid to automatically generate grid operation tickets through the techniques of big data analysis as well as neural network deep learning. CAI et al. (Xinlei et al., 2020) introduced a real-time dispatching business system based on big data applications and artificial intelligence to obtain an automated and intelligent business system. GAO et al. (Gao et al., 2019), by analyzing the power system operation and monitoring processes and incorporating artificial intelligence (AI) technology, illustrated the feasibility of forming dispatch operation tickets based on AI. Ren et al. (Ren et al., 2022) constructed a CNN-BiGRU attention-centered automatic checking model for operation tickets, which effectively improved the verification efficiency. Kumar et al. (Kumar et al., 2019; Kumar et al., 2020) proposed a novel DFOGI for element extraction and a novel HPO for GMPPT based on optimized operation of grid connected partially shaded solar photovoltaic arrays. They employed an enhanced optimal control technique based on adaptive maximization in Kalman filter (AM-MKF) to maximize the power generation of solar photovoltaic panels. Saxena et al. (Saxena et al., 2021) proposed an improved model predictive control method based on a double second-order generalized integrator for the control problem of two-stage three-phase grid connected solar photovoltaic power systems. These demonstrate the feasibility of using neural network deep learning to solve practical problems in power grids. The current generation method based on neural network deep learning is still based only on the original rules for partial modification and does not let the model use historical data or online data to perform end-to-end generation. Therefore, this paper presents a new grid equipment startup program generation method based on the new GATv2 and ERNIE models for end-to-end generation of new grid equipment startup programs through the combination of graph-based deep learning and a text generation model.

2 Basic theory of the disconnecting switch status monitoring technology based on the internet of things

Several factors need to be considered in a grid start-up program for new equipment. The first is the safety and reliability of the new equipment, which must be considered to ensure that its introduction does not pose any potential risk to the grid system as a whole. Second, compatibility with the existing system must be considered. The new equipment should be able to seamlessly integrate and work with existing equipment without causing unnecessary failures or instability. Then, efficiency and performance optimization must be considered, as the introduction of new equipment should lead to more efficient energy transfer and management, helping to improve the operational efficiency of the entire power system. New grid equipment startup programs can be constructed incorrectly for various reasons, very likely due to failures to access the new equipment after an impact on the power grid and similar factors, which can lead to grid failures and other undesirable outcomes, including accidents. In particular, the reasonableness of the site may be directly related to the safety of the personnel. Therefore, a new grid equipment startup program needs to be extremely accurate; although recommendations for new grid equipment startup programs are given in a series of professional terminology texts, as long as the equipment is established in a rigorous and correct manner, these rules can be expressed in a variety of ways. A complete new grid equipment startup program is divided into five parts: establishing the startup scope, commissioning the project, determining the scheduled startup time, setting startup conditions, and performing startup steps. The start-up scope and scheduled start-up time are often directly known, and the commissioned project is directly determined from the start-up scope; thus, new grid equipment start-up program generation is mainly based on the start-up conditions and start-up steps. Therefore, in general, compared with other types of programs, the new grid equipment startup program has differences in rigor, information asymmetry, and complexity.

- (a) Strictness of the differences: A new grid equipment startup program gives the names of the specific equipment involved, and the corresponding operations must be completely consistent with the actual grid equipment; however, for the other statements in the text, one only needs to ensure that the program is reasonable and easy to interpret and that it is in line with the actual use of the power grid as much as possible.
- (b) Information asymmetry: The topological information contained in the start-up scope, commissioned items, and scheduled start-up time of the new grid equipment start-up program is not symmetrical with the topological information contained in the start-up conditions and start-up steps.
- (c) Complexity difference: The new grid equipment startup programs for different types of equipment, different numbers of pieces of equipment, different topologies of the grid, and other complex factors are also different.

To better generate the initiation scheme, this paper uses the GATv2 model and the ERNIE-GEN model to determine the topological information and the text information, respectively,

and adopts a coding strategy based on the device type, a self-attention layer selection strategy based on the complexity of the scheme, and a device information modification strategy to better address the three abovementioned characteristics.

2.1 GATv2 model

The GATv2 model (Brody et al., 2021) is an improved model based on the graph attention network (GAT) model (Veličković et al., 2017), which solves the static attention problem in the original GAT model; i.e., the ordering of the attention weights is independent of the querying nodes, resulting in each node paying attention to the same highest-scoring neighboring node. The GATv2 model improves the expressiveness and robustness of the model by adjusting the linear transformation and the order of attention computation so that each node can focus on different neighboring nodes, thus improving the expressiveness and robustness of the model. The specific structure of the model is as follows:

Encoder: The encoder consists of the decoder portion of a multilayer unidirectional transformer, where each layer of graph convolution uses a self-attention mechanism; i.e., different weights are dynamically assigned based on the similarity and adjacency between the nodes. Thus, the information from the neighboring nodes is fused to encode the input graph structure and extract the semantic features of the nodes.

Decoder: The decoder acts as a multilayer fully connected network that can take the output vector of the encoder as input for different tasks; this network is used in this paper to obtain the node classification task output.

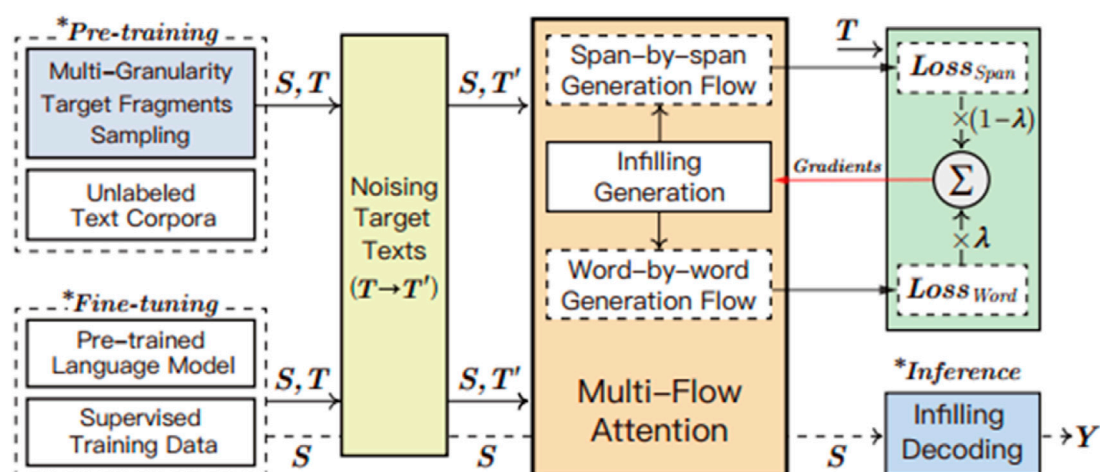
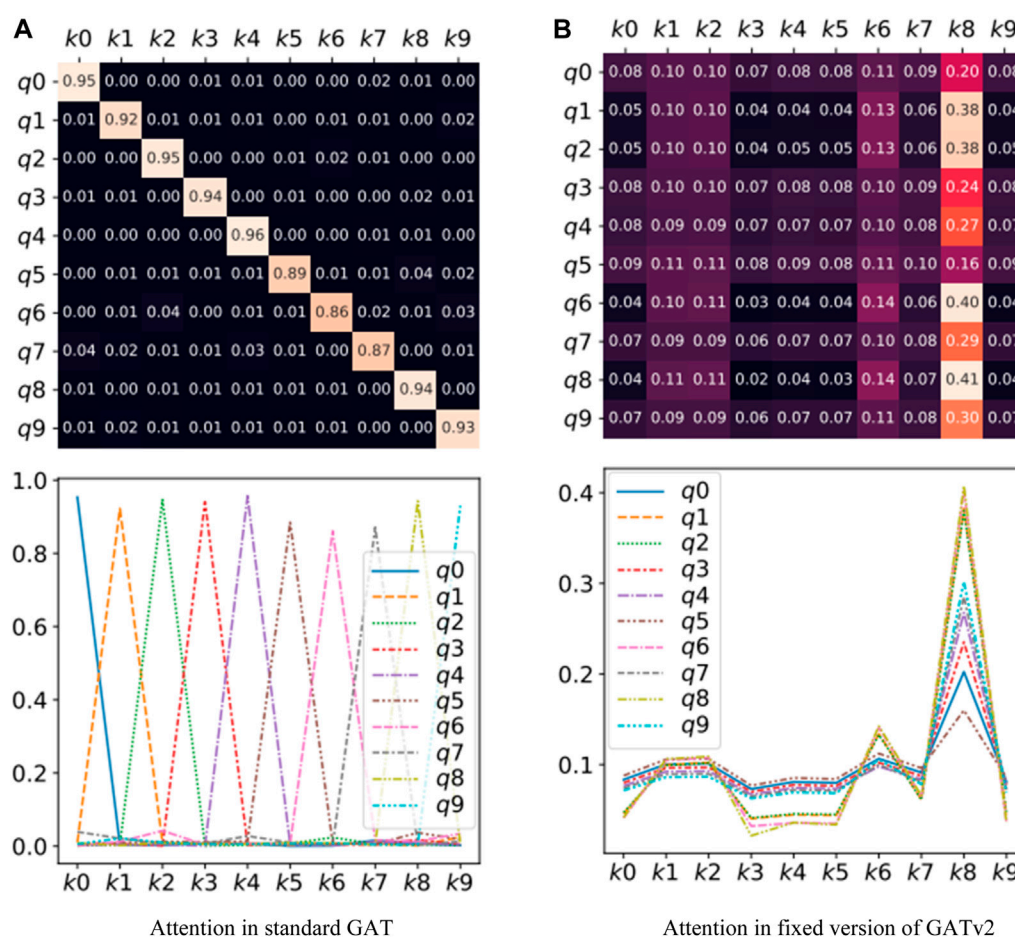
Dynamic Attention: dynamic attention is a major innovation of the GATv2 model, which realizes a dynamic attention mechanism by reversing the order of linear transformation and attention computation so that the ordering of the attention weights is affected by the query nodes, the performance of its attention mechanism compared to the original GAT model is shown in Figure 1. Where the original GAT utilizes the score function $e: R^d \times R^d \rightarrow R$ to score the nodes and the equation is shown in Formula 1:

$$e(h_i, h_j) = \text{LeakyReLU}(a^T [W h_i \| W h_j]) \quad (1)$$

where $e(h_i, h_j)$ denotes the attention weight between node i and node j , $a \in R^{2d'}$ is a learnable vector, LeakyReLU is the activation function, $W \in R^{d' \times d}$ is a learnable matrix used to vary the node features linearly, and $\|$ denotes the vector splicing operation. For node i , after calculating all the neighbor scores, use softmax to normalize the attention weight of the i th element to the j th element in the sequence and the formula is shown in Formula 2:

$$\alpha_{ij} = \text{softmax}_j(e(h_i, h_j)) = \frac{\exp(e(h_i, h_j))}{\sum_{j' \in N_i} \exp(e(h_i, h_{j'}))} \quad (2)$$

The GAT performs aggregation based on these weights to obtain the hidden features of the i th node in the GAT and the formula is shown in Formula 3:



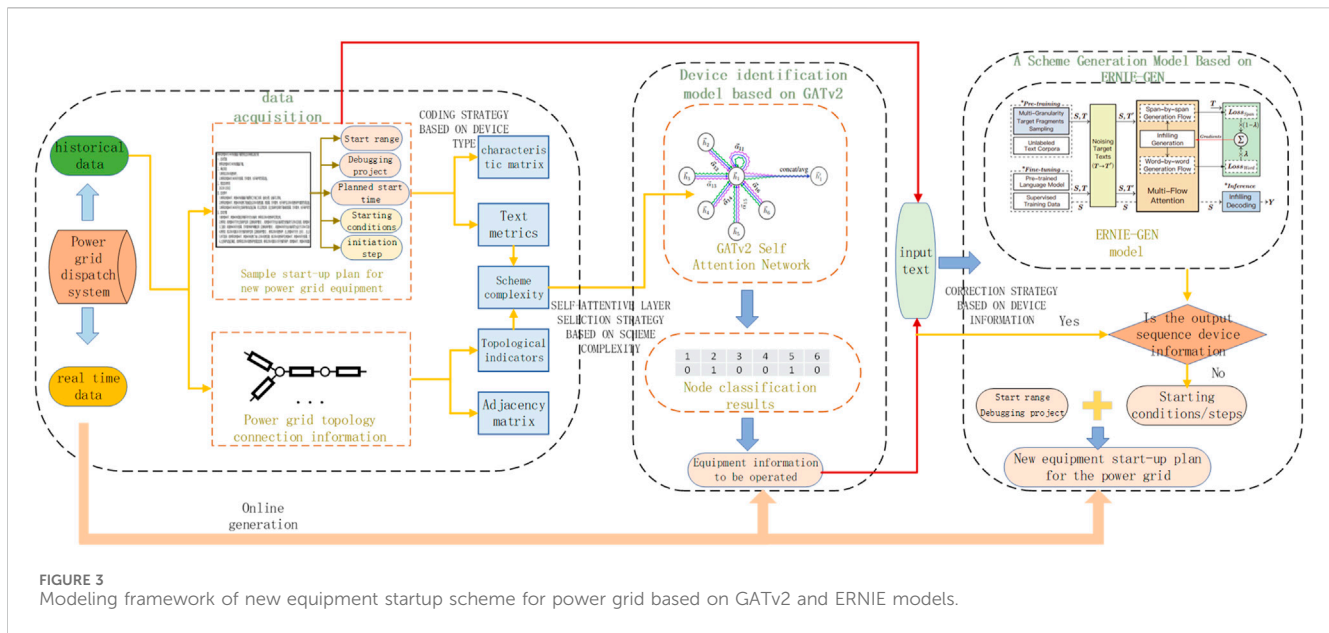


FIGURE 3
Modeling framework of new equipment startup scheme for power grid based on GATv2 and ERNIE models.

$$h'_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} W h_j \right) \quad (3)$$

The dynamic attention of GATv2 solves the problem of the standard GAT problem in which W , a are used sequentially in the scoring function, resulting in a linear layer that collapses into a single layer at the end. By moving an out of the nonlinear result and then running it, in addition, the query-key pair is concatenated first, and then W is used to perform the linear transformation, i.e., at this time, the expression of $e(h_i, h_j)$ is shown in Formula 4:

$$e(h_i, h_j) = a^T \text{LeakyReLU}(W[h_i \| h_j]) \quad (4)$$

2.2 GATv2 model

ERNIE-GEN is a generative pre-training model proposed by Baidu and the schematic diagram of its framework is shown in Figure 2, ERNIE-GEN model is a generative pre-training model, which is based on Transformer's seq2seq framework, targeting the exposure bias problem in training and the insufficient interaction between encoder and decoder in pre-training that is the exposure bias problem and the equations are shown in Formula 5, 6:

$$\text{Training: } y_{i+1} \leftarrow MH - \text{Attn}(Q = t_i, KV = [S, t \leq i]) \quad (5)$$

$$\text{Decoding: } y_{i+1} \leftarrow MH - \text{Attn}(Q = y_i, KV = [S, y \leq i]) \quad (6)$$

Where y and t denote the predicted character vector and ground truth character vector, respectively; S is the representation of the encoder side; and the direct influence on the prediction of y_{i+1} is represented as Q , which is used to converge the above representations and is also where the difference between the training and decoding phases has the most direct effect. At the same time, KV differs in training and decoding, but the effect on y_{i+1} is weaker. To solve this problem, ERNIE-GEN introduces several innovative mechanisms to improve the quality and

efficiency of natural language generation. The main features of the ERNIE-GEN model are as follows:

Multiflow Attention: Multiple distinct attention flows are added between the encoder and decoder to enhance the codec interaction in such a way that integrated word-by-word and span-by-span generation flows can be applied in parallel with the shared context flow. The multistream computational power's equations are shown in Formula 7–9:

$$X^{(l+1)} \leftarrow MH - \text{Attn}(Q = X^{(l)}, KV = X^{(l)}, M_c) \quad (7)$$

$$A_W^{(l+1)} \leftarrow MH - \text{Attn}(Q = A_W^{(l)}, KV = [X^{(l)}, A_W^{(l)}], M_w) \quad (8)$$

$$A_S^{(l+1)} \leftarrow MH - \text{Attn}(Q = A_S^{(l)}, KV = [X^{(l)}, A_S^{(l)}], M_s) \quad (9)$$

where X is the connection between S and T' , $X(l)$ is the layer- l vector sequence of the context stream, and $A_W(l)$ and $A_S(l)$ are the layer- l vector sequences of the verbatim generation stream and the word-by-word generation stream, respectively.

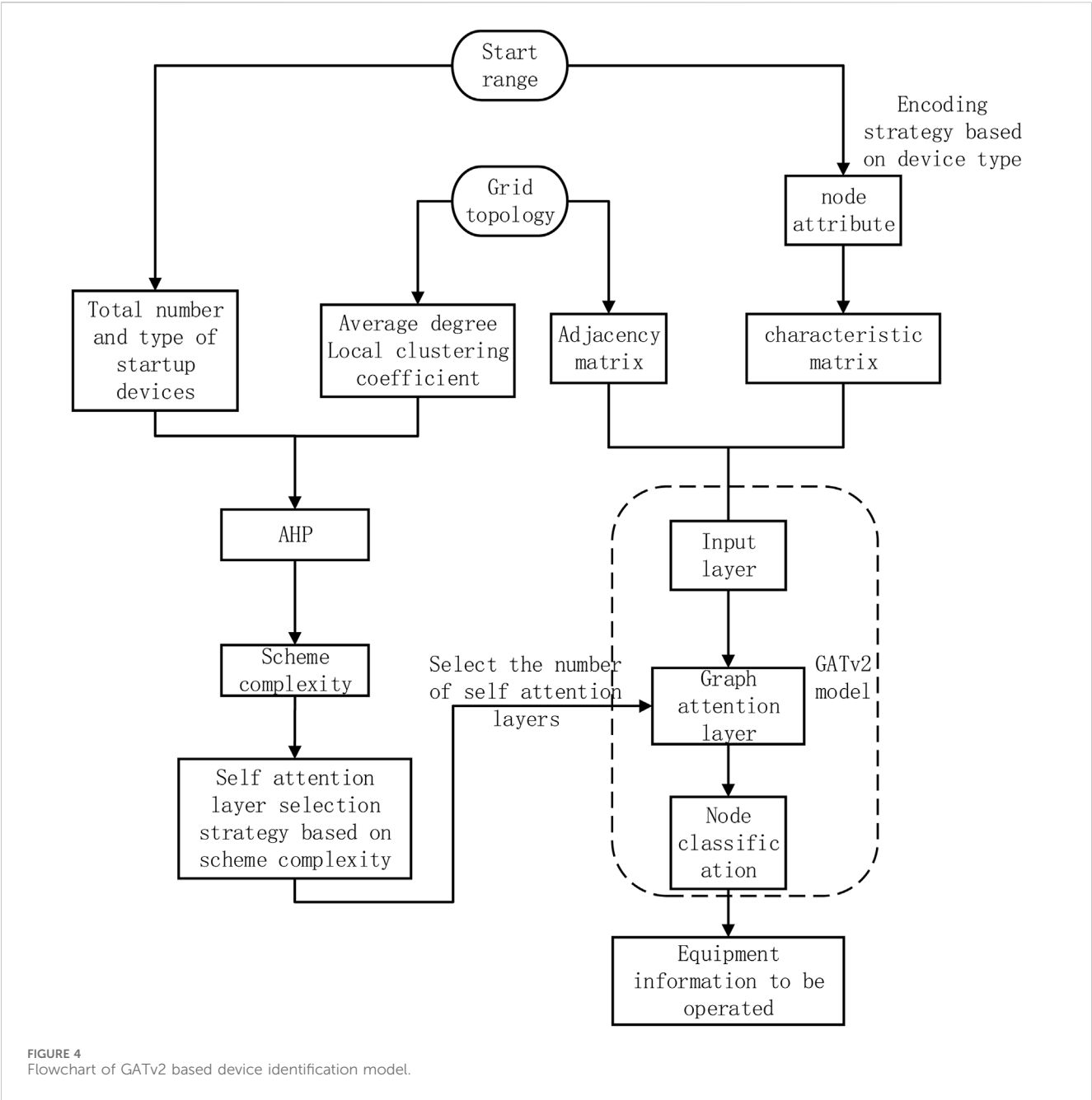
Infilling Generation Mechanism: By adding a special symbol [ATTN] after each character in the decoder, the model's attention is shifted from the last word to all previous representations to attenuate the dependence on the previous character while unifying the training and decoding conditions.

Noise-Aware Generation Method (NGM): Random noise is added to the input sequence of the decoder during training to train the model to perceive errors and attenuate the impact of errors on subsequent generation by adjusting the attention weights.

Span-by-Span Learning Paradigm: Instead of predicting only one character at each step during training, a semantically complete segment is predicted to improve the coherence and accuracy of generation.

3 Construction of a GATv2 and ERNIE-GEN based model for generating new equipment startup scenarios for the grid

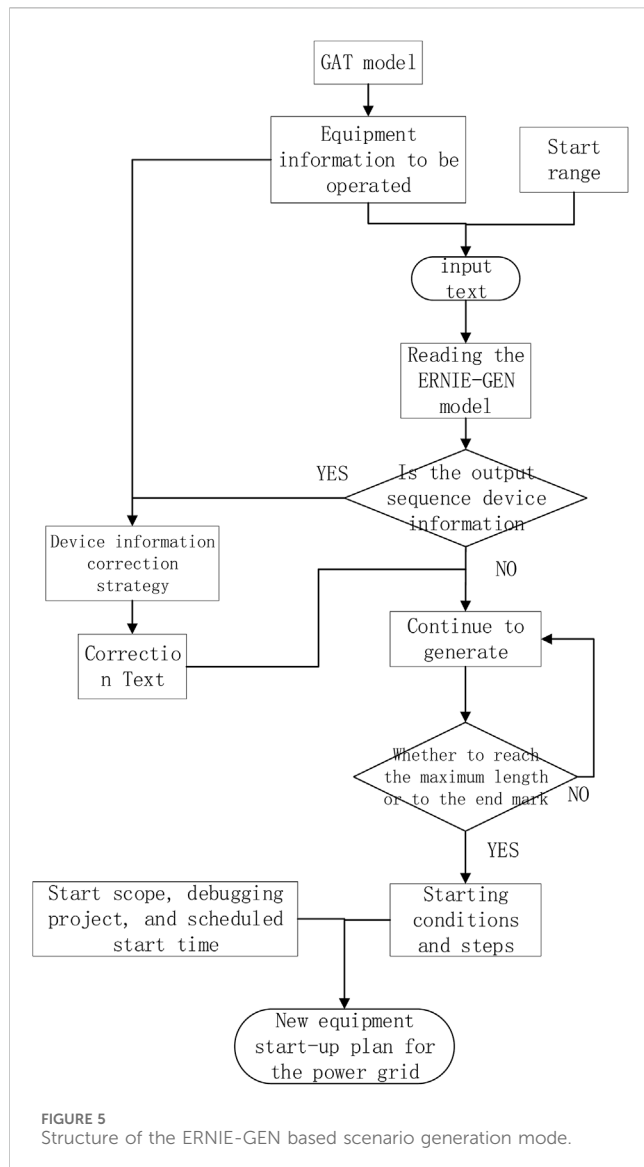
In this paper, we use Python for new grid device bootstrap scheme generation and use the PyCharm platform to realize end-to-



end bootstrap scheme generation. According to the above discussion, new grid device initiation schemes have rigor variability, information asymmetry, and complexity variability. In this paper, we propose a fusion topology for a new grid device initiation scheme based on the GATv2 model as well as the ERNIE-GEN model. Additionally, we propose a coding strategy based on device type, a self-attention layer selection strategy based on scheme complexity, and a correction strategy based on device information to address these three characteristics in the generation of new device initiation schemes for power grids.

The model first inputs the startup range into the GATv2 model after applying the device type-based coding

strategy, obtains all the information on the devices to be operated based on the self-attention layer selection strategy according to the scheme complexity and then passes this information to the trained ERNIE-GEN text generation model. The ERNIE-GEN text generation model takes the startup range and all the device information obtained from GATv2 as text input and obtains the startup conditions and startup steps based on the device information modification strategy. Finally, the output is directly combined with the known startup scope, scheduled startup time, and commissioned items to obtain the final new grid equipment startup program. The framework structure is shown in Figure 3.



3.1 GATv2-based device identification model construction

The GATv2-based device identification model obtains node attributes through the startup range combined with the topology information of the device type-based coding strategy and inputs them as feature matrices. The adjacency matrix obtained based on the grid topology is also fed as input to the GATv2 model, and node classification is used to obtain information about the devices to address the information asymmetry of the grid's new device activation scheme. To address the complexity variability of the startup scheme, the model determines the number of layers of the graph self-attention layer through a selection strategy based on the scheme complexity. The structure of the model is schematically shown in Figure 4.

3.1.1 Coding strategy based on device type

To allow the model to better recognize device types so that it can better handle grid topology information, this paper adopts a coding strategy based on device types. In this paper, all device types are

treated as nodes, and devices of the same type are assigned the same encoding as a node attribute instead of directly encoding the name of the device as a node attribute. This approach allows the model to read the device information more efficiently, thus avoiding interference caused by unnecessary information.

3.1.2 Self-attentive layer selection strategy based on scheme complexity

To prevent the overfitting and underfitting of the model due to the complexity variability of the startup scheme, this paper presents a self-attentive hierarchical selection strategy based on the scheme's complexity. By analyzing the characteristics of the new grid equipment startup scheme and incorporating the relevant literature, this paper uses the four indicators of the total number of startup devices, the number of types, the average degree, and the local clustering coefficient based on the hierarchical analysis method to measure the scheme complexity.

Total number of devices to be started: the total number of devices to be activated in the grid's new device startup program;

Number of types: the number of types of equipment to be activated in the grid's new equipment activation program;

Average degree: the average of the degrees of all nodes in the grid topology;

Local clustering coefficient: the ratio of the number of edges that are connected between neighboring nodes of the devices that need to be activated to the maximum possible number of connections. In this paper, we take the maximum value among all the activated devices.

For the hierarchical analysis method, this paper adopts the expert scoring method to determine the relative importance of two factors and constructs its judgment matrix A , which is shown in Formula 10.

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \quad (10)$$

Where, a_{ij} denotes the importance of factor i relative to factor j , $a_{ii} = 1$, $a_{ij} = 1/a_{ji}$, after obtaining the judgment matrix.

Formula 11 is applied using the maximum eigenvalue and eigenvector to solve for the weight vector and the consistency index:

$$AW = \lambda_{\max} W \quad (11)$$

where A is the judgment matrix, W is the eigenvector, and λ_{\max} is the maximum eigenvalue. Each component of W represents the weight of a factor, and these values need to be normalized. λ_{\max} represents the degree of consistency of the judgment matrix; if $\lambda_{\max} = n$, then the judgment matrix is completely consistent. If $\lambda_{\max} > n$, there is some inconsistency in the judgment matrix; therefore, the formula for verifying the consistency of the judgment matrix is as shown in Eqs 12, 13:

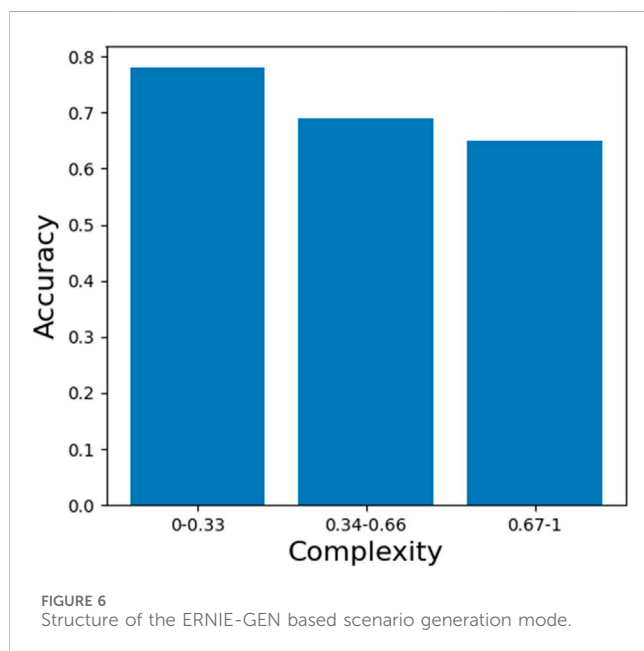
$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (12)$$

$$CR = \frac{CI}{RI} \quad (13)$$

where CI is the consistency index, CR is the consistency ratio and RI is the random consistency index, which is obtained based on the

TABLE 1 Data distribution table for new device startup plan.

	Type of line protection	Terminal box replacement type	CT replacement type	Line start type (shock)
quantities	30	100	100	40
Text length	547–937	547–1531	599–1642	430–472
Maximum length of a single sentence	91	117	115	72
Maximum number of device types	3	4	4	2



average of the random matrix. If $CR < 0.1$, it means that the consistency of the judgment matrix is acceptable; otherwise, the judgment matrix needs to be modified. Finally, its complexity $scplex$ is calculated using the formula shown in Eq. 14:

$$scplex = W_1 * num + W_2 * type + W_3 * avdg + W_4 * cluster \quad (14)$$

where num is the total number of startup devices, $type$ is the number of types, $avdg$ is the average degree, $cluster$ is the local clustering coefficient, and W_1 , W_2 , W_3 , and W_4 are the respective coefficients. Finally, the complexity is divided into three types, simple, more complex and complex, and the number of layers of self-attention increases according to the increase in complexity.

3.2 Construction of ERNIE-GEN based scenario generation models

The ERNIE-GEN-based (Xiao et al., 2020) scenario generates the model by inputting the input text as well as the corresponding output text into the model, which is trained. The text is fine-tuned through preprocessing, encoding and decoding, and text generation;

TABLE 2 Correspondence between program complexity and number of self-attention layers.

	Program complexity		
	0–0.33	0.34–0.66	0.67–1
Number of layers of self-attention	1	2	3

based on whether the loss function converges, it is determined whether to make parameter updates.

When the model training is complete, the program generation model will receive the information about the equipment to be operated, which is generated by the GATv2-based equipment recognition model combined with the startup scope as the input text, and the equipment information is combined with the correction strategy to generate the startup conditions and startup steps in the startup program. The generated startup conditions and steps, together with the known startup scope, commissioned items, and scheduled startup time, directly form the startup plan for the new equipment in the grid. The structure of the model is schematically shown in Figure 5.

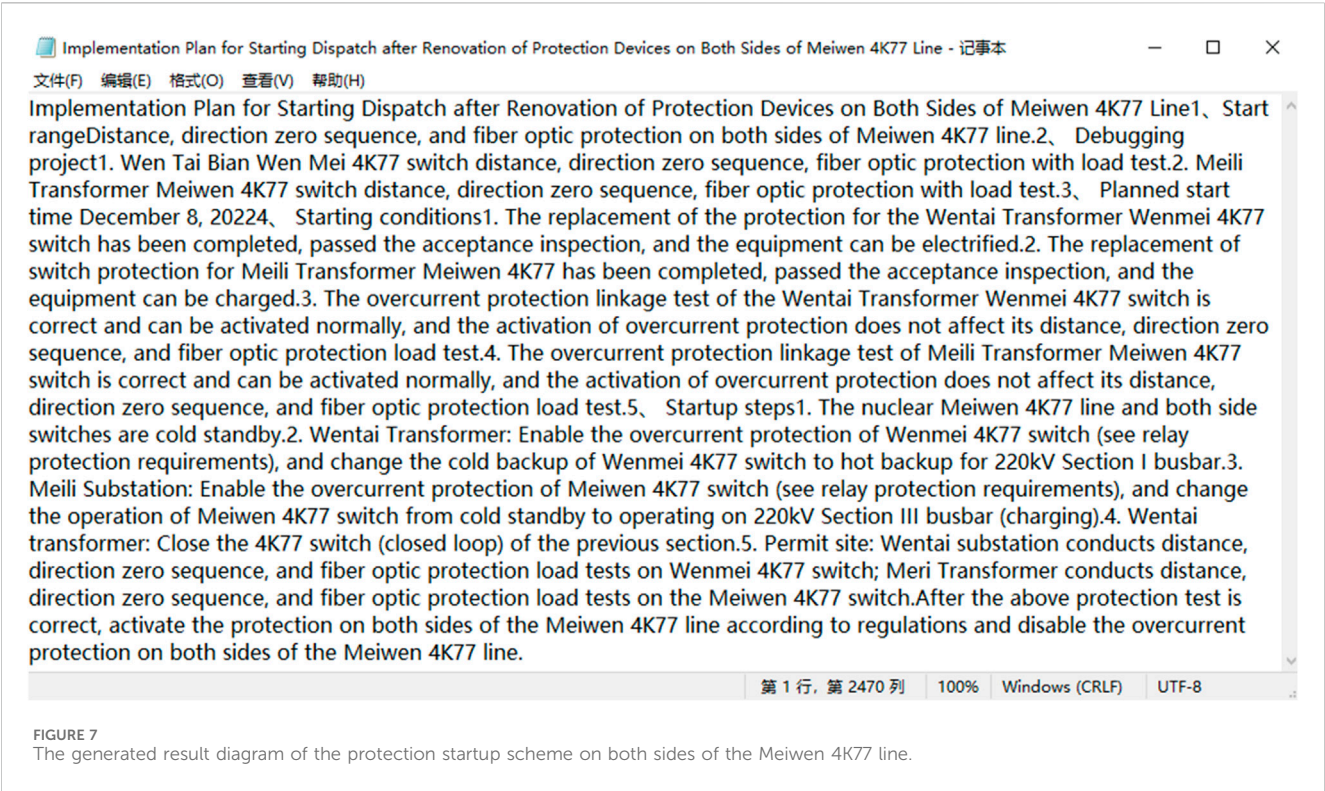
3.2.1 Correction strategy based on device information

The equipment information in the new equipment startup program of the power grid is rigorous, but when generating text based on the equipment information, it is possible that the generation process is not completely rigorous. Therefore, this paper presents a correction strategy based on device information. When the sequence of the generated text output is the device information, it will be directly selected from the information on the devices to be activated, after which the text will be corrected.

4 Case validation

4.1 Case data and simulation

In this paper, according to the actual grid history of a local new grid equipment startup program that is to be verified by offline calculations, the startup program-specific data include the line protection type; terminal box replacement type; CT replacement type; and line startup type (impact), of which there are four. These data are shown in Table 1.



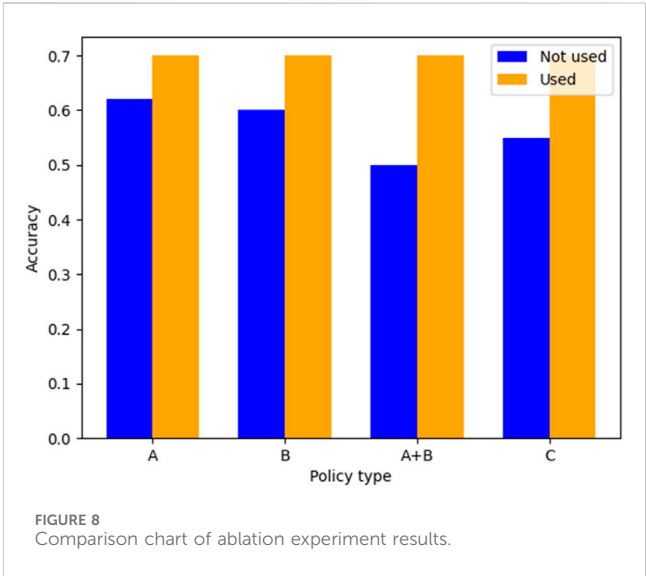
According to each specific startup scenario, the startup range is constructed according to the method in Section 3 based on the device identification model of GATv2 after verifying that it corresponds to the grid topology, and the correspondence between the scenario complexity and the self-attention layer is shown in Table 2.

The samples under each scenario complexity are partitioned at a ratio of 8:2, and the accuracy of the final model obtained for each scenario complexity case is shown in Figure 6.

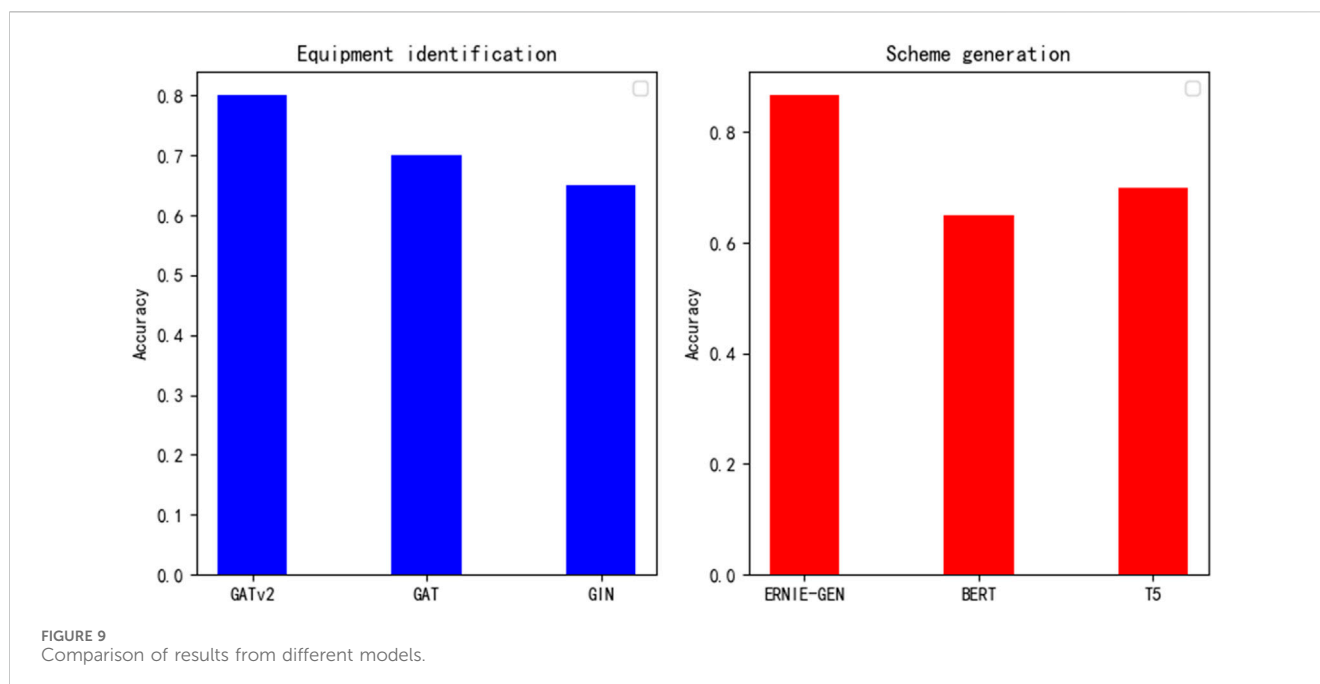
The information of the equipment to be operated and the range of activation in the activation scheme are taken as input texts, and the activation conditions and steps are taken as the corresponding output texts, and they are corresponded to each other to generate the model training and reading according to the scheme of ERNIE-GEN in Section 3. Since the non-equipment words in the text of the grid new equipment startup program only need to be correct and fluent, there is no strict requirement as in the case of equipment. If the direct use of text similarity detection is difficult to detect whether the semantics of non-equipment vocabulary change, so the generated text is uploaded to the traditional five-proof inspection system for inspection as the accuracy, and finally obtained the accuracy of ERNIE-GEN's program generation model is 0.876,545. One of the new device startup schemes obtained is shown in Figure 7.

4.2 Verification of ablation experiments

To better demonstrate the advantages of the model in addition to the new device startup scheme generation method, this paper focuses on ablation experiments on three strategies, namely, the device type-based coding strategy, the scheme complexity-based self-attention layer selection strategy,



and the device information-based correction strategy. For the sake of narrative convenience, the device type-based coding strategy module is denoted as A, the scheme complexity-based self-attention layer selection strategy is denoted as B, and the device information-based correction strategy is denoted as C. Among them, the GATv2-based device recognition model is evaluated using the accuracy rate as the index, and the average accuracy rate of the three different complexity levels is taken as the most useful index; the ERNIE-GEN-based scheme generation model is evaluated using the device name accuracy rate as the accuracy rate. The results are shown in Figure 8.



By analyzing the comparative graphs of the ablation experiment results, we can see that all three strategies achieve better results under the corresponding models. Combining the three characteristics of the new grid equipment startup scheme—strictness variability, information asymmetry, and complexity variability—we can see that the three strategies achieve better performance on the GATv2 model as well as on the ERNIE-GEN model and generate a strict and reasonable new equipment startup scheme.

4.3 Comparison with other models

To further demonstrate the advantages of this model in addition to new grid equipment startup program generation, this paper uses the GAT and GIN models for comparison. Because the GIN model does not have a self-attention layer, only the GAT and GATv2 models are compared using this strategy; similarly, the BERT and T5 models are used to compare the startup scheme generation models, and the results obtained for both are shown in Figure 9.

From the model result comparison diagram, we can easily see that in terms of the device identification model and the generation model, among the models compared in this paper, the proposed model achieved the highest accuracy. Combined with the extremely high accuracy requirements for generating a new grid equipment startup scheme, the fusion topology of the new grid equipment startup scheme in this paper ensures the accuracy of the scheme.

5 Conclusion

With the rapid progress of science and technology, the continuous growth of energy demand, the rapid construction

of power grids, and high-speed development, a large number of new equipment requires access to power grids. The rationality of a new grid equipment startup program is related to the safety of the grid as well as the safety of personnel; thus, many countries use new grid equipment programs as grid operation tickets for evaluation.

At present, power grid equipment startup program generation methods are divided into rule-based generation methods and neural network-based deep learning generation methods. Rule-based methods have a certain degree of effectiveness in the regions where they were established, but in general, the rules often need to be revised during migration to other regions. Current generation methods based on neural network deep learning generation, in general, have not been able to avoid this limitation either. Therefore, in this paper, from a practical point of view, we fully consider the difference in rigor, information asymmetry, and difference in complexity of new equipment startup programs for power grids; use the GATv2 model and the ERNIE-GEN model to address the topology and the text, respectively; and fully utilize a coding strategy based on the type of equipment, a self-attention layer selection strategy based on the complexity of the program, and a revision strategy based on the information of the equipment to address the above three characteristics. The model is also verified to be able to efficiently and accurately generate new equipment startup schemes for power grids through actual power grids, ablation experiments are used to verify the efficiency of the three strategies, and model comparison experiments are used to verify the accuracy of the proposed model. The results show that the method proposed in this paper can efficiently and accurately generate new grid equipment startup schemes. This approach provides a new idea for the development of new grid equipment startup schemes through deep learning.

Since the method of this paper uses an offline dataset for training that does not cover all types of new equipment startup scenarios for power grids, When the example model in this article starts scheduling

for new types of equipment, such as 220 KV equipment in a certain location, the generated startup plan has a certain deviation from the actual startup party, its training needs to be enhanced for use in real power grids to better utilize the requirements of real power grids for new equipment startup scenario generation.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

TM: Writing—original draft. XL: Writing—original draft. XW: Writing—original draft. LW: Writing—original draft, Writing—review and editing. WT: Writing—original draft, Writing—review and editing. LS: Writing—original draft, Writing—review and editing. XG: Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work

was supported by the Science and Technology Project of State Grid Shanxi Electric Power Company (No: 520530230006).

Conflict of interest

Authors TM, XL, and XiW were employed by State Grid Shanxi Electric Power Company.

Authors LS and XG were employed by Beijing Kedong Electric Power Control System.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from State Grid Shanxi Electric Power Company. The funder had the following involvement in the study: the data collection and analysis part.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Ans, Z., Yi, M., Liu, J., Peng, Z., Yu, S., et al. (2021). Intelligent filling method of power grid working ticket based on historical ticket knowledge base. *Front. Energy Res.* 9, 813855. doi:10.3389/fenrg.2021.813855
- Brody, S., Alon, U., and Yahav, E. (2021). How attentive are graph attention networks? *arXiv Prepr. arXiv:2105.14491*. doi:10.48550/arXiv.2105.14491
- Cai, X., and Qi, Y. (2021). Discussion on grid auxiliary control system based on big data application and artificial intelligence decision-making. *Electrotechnology* (06), 40–42. doi:10.19768/j.cnki.dgjs.2021.06.016
- Chen, S., Song, S., Lanxin, L. I., et al. (2009). An overview of smart grid technology. *Grid Technol.* 33 (08), 1–7.
- Dong, C., Zhao, J., Wen, F., et al. (2014). From smart grid to energy internet: basic concepts and research framework. *Power Syst. Autom.* 38 (15), 1–11. doi:10.7500/AEPS20140613007
- Gao, D. W., Wang, Q., Zhang, F., Yang, X., Huang, Z., Ma, S., et al. (2019). Application of AI techniques in monitoring and operation of power systems. *Front. Energy* 13, 71–85. doi:10.1007/s11708-018-0589-4
- Gong, J., Sun, H., Tang, L., et al. (2006). Scheduling operation ticket expert system based on network analysis and intelligent reasoning. *Grid Technol.* (17), 19–24. doi:10.3321/j.issn:1000-3673.2006.17.004
- Guo, X., Zhu, C., Cao, J., et al. (2006). Comprehensive error prevention strategy and intelligent operation ticket generation system based on multiple intelligences. *Grid Technol.* (04), 50–54. doi:10.13335/j.1000-3673.pst.2006.04.010
- Harbor, WANG, Xiaofei, H. U., Cao, K., et al. (2014). Risk evaluation method for new equipment startup program in power grid. *Power Syst. Prot. Control* 42 (17), 66–71.
- Hu, H., Sun, H., Zhang, B., et al. (2002). Research and application of substation operation ticket expert system. *Power Autom. Equip.* (08), 42–45. doi:10.3969/j.issn.1006-6047.2002.08.013
- Jiang, H., Jie, L. I. U., Zhaojun, LIANG, et al. (2005). Automatic generation system of substation operation ticket based on multi-agent. *New Technol. Electr. Eng. Electr.* 2005 (04), 18–21+67. doi:10.3969/j.issn.1003-3076.2005.04.005
- Kumar, N., Singh, B., and Panigrahi, B. K. (2019). Grid synchronisation framework for partially shaded solar PV-based microgrid using intelligent control strategy. *IET generation, Transm. distribution* 13 (6), 829–837. doi:10.1049/iet-gtd.2018.6079
- Kumar, N., Singh, B., Wang, J., and Panigrahi, B. K. (2020). A framework of L-HC and AM-MKF for accurate harmonic supportive control schemes. *IEEE Trans. Circuits Syst. I Regul. Pap.* 67 (12), 5246–5256. doi:10.1109/tcsi.2020.2996775
- Li, M., Wang, Y., Li, X., et al. (2010). Analysis of intelligent substation and technical characteristics. *Power Syst. Prot. Control* 38 (18), 59–62.
- Li, Y., Liu, Y., Wang, X., et al. (2016). Design and realization of intelligent programming system for grid new equipment startup program based on MAS. *Power Syst. Prot. Control* 44 (12), 119–124.
- Liu, K., Sheng, W., Li, Z., Liu, F., Liu, Q., Huang, Y., et al. (2023). An energy optimal schedule method for distribution network considering the access of distributed generation and energy storage. *IET Generation, Transm. Distribution* 17 (13), 2996–3015. doi:10.1049/gtd.12855
- Liu, Q., Zhou, M., Gengyin, L. I., et al. (2005). Grid dispatching operation ticket system with computational analysis function. *Grid Technol.* (07), 68–73. doi:10.3321/j.issn:1000-3673.2005.07.015
- Mengchao, L. I., Wang, Y., Xianwei, L. I., et al. (2010). Analysis of intelligent substation and technical characteristics. *Power Syst. Prot. Control* 38 (18), 59–62+79. doi:10.3969/j.issn.1674-3415.2010.18.012
- Otuozee, A. O., Mustafa, M. W., and Larik, R. M. (2018). Smart grids security challenges: classification by sources of threats. *J. Electr. Syst. Inf. Technol.* 5 (3), 468–483. doi:10.1016/j.jesit.2018.01.001
- Ren, S., Lv, X., Sun, K., et al. (2022). “Chinese operation ticket automatic verification method based on CNN-BiGRU-Attention in intelligent substations,” in 2022 IEEE International Conference on High Voltage Engineering and Applications (ICHVE), USA, 25–29 Sept. 2022 (IEEE), 1–4.
- Ren, C., Niu, S., Xianbo, K. E., et al. (2022). Research on intelligent programming system for new equipment startup program in power grid. *Electr. Meas. Instrum.* 59 (07), 78–84. doi:10.19753/j.issn1001-1390.2022.07.011
- Sanjab, A., Saad, W., Guvenc, I., et al. (2016). Smart grid security: threats, challenges, and solutions. *arXiv Prepr. arXiv:1606.06992*. doi:10.48550/arXiv.1606.06992
- Saxena, V., Kumar, N., Singh, B., and Panigrahi, B. K. (2021). An MPC based algorithm for a multipurpose grid integrated solar PV system with enhanced power quality and PCC voltage assist. *IEEE Trans. Energy Convers.* 36 (2), 1469–1478. doi:10.1109/tec.2021.3059754

- Song, H. (1999). Grid dispatching operation ticket management expert system and application. *Power Syst. Autom.* (19), 58–59. doi:10.3321/j.issn:1000-1026.1999.19.013
- Sun, C. C., Hahn, A., and Liu, C. C. (2018). Cyber security of a power grid: state-of-the-art. *Int. J. Electr. Power and Energy Syst.* 99, 45–56. doi:10.1016/j.ijepes.2017.12.020
- Tang, L., Zhang, B., Sun, H., et al. (2001). A generalized cognitive model in grid operation ticket expert system. *Power Syst. Autom.* (22), 6–9+21. doi:10.3321/j.issn:1000-1026.2001.22.002
- Veličković, P., Cucurull, G., Casanova, A., et al. (2017). Graph attention networks. *arXiv Prepr. arXiv:1710.10903*. doi:10.48550/arXiv.1710.10903
- Wang, H., Hu, X., Cao, K., et al. (2014). Risk evaluation method for new equipment startup program in power grid. *Power Syst. Prot. Control* 42 (17), 66–71.
- Wei, J., Wang, S., Huang, S., Peng, H., Wei, L., et al. (2023). Optimal post-fault recovery control based on topology reconfiguration of wind farm collection systems. *IEEE Trans. Industry Appl.* 60, 3664–3675. doi:10.1109/tia.2023.3344059
- Wu, Q., Liu, W., and Yang, Y. (2006). Development and realization of intelligent grid scheduling decision support system. *Power Syst. Autom.* (12), 79–83. doi:10.3321/j.issn:1000-1026.2006.12.017
- Xiao, D., Zhang, H., Li, Y., Sun, Y., Tian, H., Wu, H., et al. (2020). Ernie-gen: an enhanced multi-flow pre-training and fine-tuning framework for natural language generation. *arXiv Prepr. arXiv:2001.11314*. doi:10.24963/ijcai.2020/553
- Xinlei, C. A. I., Rongfu, Q. I. U., Yanli, C. U. I., et al. (2020). “Power grid auxiliary control system based on big data application and artificial intelligence decision,” in 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), USA, 23–25 Oct. 2020 (IEEE), 154–157.
- Yang, J., Ming, H. U., Qiong, W. U., et al. (2004). Design and development of an expert system for operating tickets in power grid dispatching. *Relay* (15), 45–47+58.
- Yi, Y., and Wenpeng, L. (2009). Review of Smart Grid. *Chin. J. Electr. Eng.* 29 (37), 1–8.
- Yonggang, L. I., Yan, L., Xinming, W., et al. (2016). Design and realization of intelligent programming system for grid new equipment startup program based on MAS. *Power Syst. Prot. Control* 44 (12), 119–124. doi:10.7667/PSPC151214
- Yu, Y., and Luan, W. (2009). Review of smart grid. *China J. Electr. Eng.* 29 (34), 1–8.
- Yuan, F., Feng, X., Li, W., et al. (2022). “Research and development of an improved intelligent operation ticket system,” in 2022 International Conference on Wireless Communications, Electrical Engineering and Automation (WCEEA), USA, 15–16 Oct. 2022 (IEEE), 210–213.
- Zhang, W., Liu, Z., Mingjun, W., et al. (2009). Research progress and development trend of smart grid. *Grid Technol.* 33 (13), 1–11.
- Zhou, M., Ren, J., Gengyin, L. I., et al. (2004). Research and realization of grid scheduling operation ticket guidance system based on multi-intelligence. *China J. Electr. Eng.* (04), 62–66+213. doi:10.3321/j.issn:0258-8013.2004.04.011
- Zhou, M., and Yang, G. (2004). Automatic generation and management system of operation tickets for new smart grid dispatch. *Power Syst. Autom.* (11), 71–74. doi:10.3321/j.issn:1000-1026.2004.11.017
- Zhu, B., Wu, H., Tong, C., et al. (2022). Research on automatic generation system of grid scheduling operation ticket based on expert system. *Electron. Devices* 45 (04), 925–930. doi:10.3969/j.issn.1005-9490.2022.04.027
- Zhu, Y., Zhang, J., Yang, Z., et al. (2003). Object-oriented generalized expert system for work ticket and operation ticket in agricultural substation. *Grid Technol.* (04), 27–30. doi:10.3321/j.issn:1000-3673.2003.04.006



OPEN ACCESS

EDITED BY

Fuqi Ma,
Xi'an University of Technology, China

REVIEWED BY

Quan Zhou,
University of Birmingham, United Kingdom
Chongfeng Wei,
Queen's University Belfast, United Kingdom

*CORRESPONDENCE

Qikun Chen,
✉ qikun.chen@glasgow.ac.uk

RECEIVED 17 November 2023

ACCEPTED 03 May 2024

PUBLISHED 14 June 2024

CITATION

Jiao F, Ma Z, Chen Q, Zhang F and Zhao D (2024), A causal reasoning approach for power transformer failure diagnosis. *Front. Energy Res.* 12:1340421. doi: 10.3389/fenrg.2024.1340421

COPYRIGHT

© 2024 Jiao, Ma, Chen, Zhang and Zhao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A causal reasoning approach for power transformer failure diagnosis

Fei Jiao¹, Zhenyuan Ma¹, Qikun Chen^{1*}, Fengda Zhang² and Dezong Zhao¹

¹China Electric Power Research Institute, Department of Artificial Intelligence Application, Beijing, China, ²State Grid Shandong Electric Power Research Institute, Jinan, China

Extensive research validates the effectiveness of employing Dissolved Gas Analysis (DGA) for diagnosing electric power transformer failures. However, a significant portion of existing research focuses on static data for classifying failure types, lacking a thorough exploration of causality. This study proposes an approach integrating causality and the DGA framework to infer power transformer failures. Validation through 96 historical samples from diverse transformers demonstrates the capability of this method to identify probable abnormal failures of the power transformer accurately. The proposed causal reasoning method is able to diagnose all common transformer states, accounting for the level of severity in both electrical and thermal failures, and with an accuracy of 95.8%.

KEYWORDS

failure diagnosis, dissolved gas analysis, causality, ratio-based DGA, Bayesian network (BN)

1 Introduction

In modern power systems, power transformers are vital for uninterrupted energy supply. The uninterrupted functioning of various industries and daily life is contingent upon the reliable transmission and distribution facilitated by power transformers. Early failure detection is essential to prevent disruptions and ensure the stability and resilience of the power system.

Diagnosing failures in power transformers is a complex task, and Dissolved Gas Analysis (DGA) is one of the most effective methods for this purpose. DGA evaluates transformer faults by measuring concentrations of gases in transformer oil, such as hydrogen (H₂), methane (CH₄), ethane (C₂H₆), ethylene (C₂H₄), acetylene (C₂H₂), carbon monoxide (CO), and carbon dioxide (CO₂) (Kari et al., 2018). These gases can indicate different types of faults. Partial discharges generate hydrogen and methane, and the presence of acetylene and hydrogen can confirm arcing (Gouda et al., 2016). The analysis of these gases, including their levels, and proportions provides crucial insights for identifying potential malfunctions in transformers.

Various methods are used to interpret DGA data, while the conventional approaches are ratio methods like the Doernenburg ratio, Rogers ratio, and the Duval triangle (Aizpurua et al., 2018). However, while these techniques are highly accurate, they are time-consuming and expensive (Aizpurua et al., 2018). To address these challenges, there has been a shift towards intelligent diagnostic approaches that integrate soft computing methods with DGA for cost reduction (Tomsovic et al., 1993; Guo et al., 2019). For instance,

the Dempster-Shafer Theory demonstrates that soft computing methods can achieve high accuracy with large sample databases (Min and Chang, 2009). However, the adoption of machine learning in transformer failure diagnosis is hampered by the scarcity of failure data caused by the frequency of transformer failures.

To address the challenge of low transparency and the lacking of training data, this study proposes a method that integrates the DGA with causal reasoning to diagnose the transformer failures. In general, failure diagnosis and explanation are rooted in the fundamental concept of causality Lewis (1986). Achieving a clear understanding of these processes relies on accurately identifying the interdependence and causal relationships among their constituent components, as discussed in further research Özgür-Ünlüakın et al. (2021). Causal reasoning, an emerging approach, leverages knowledge of causal relationships between variables to predict outcomes based on cause-and-effect models Pearl and Mackenzie (2018). It effectively reduces the requirements for training datasets as it leverages knowledge of causal relationships between dissolved gases and equipment failures instead of static data regression.

The paper is organized as follows. After the introduction in Section 1, the conventional failure diagnosis methods are presented in Section 2. The methodology of the proposed system is proposed in section 3. The experimental results including the comparison with three conventional methods and discussions are stated in section 4. Finally, the conclusions are summarized in section 5.

2 Failure diagnosis

The generation of the hydrocarbon gases within the normal operating transformers would be caused by thermal and electrical stresses, which can indicate potential problems within the transformer (Syafuruddin and Nugroho, 2020). As transformers age, some gas generation is anticipated; distinguishing between normal and excessive gassing rates is crucial. Normal gas generation in transformers is influenced by several factors. These include transformer design, loading, and the type of insulating material used. To identify abnormal behavior in transformers, standard gassing rates are employed as a universal metric (Ali et al., 2023).

The breakdown of insulation in transformers leads to the production of crucial gases such as H_2 , CH_4 , C_2H_6 , C_2H_4 , C_2H_2 . Other gases such as CO and oxygen (O_2) are also present, originating from the degradation of cellulose insulation (Kari et al., 2018). Moreover, external factors like CO_2 , nitrogen (N_2), and moisture can be absorbed from the air due to an oil/air interface or tank leak. The specific gases generated and their quantities depend on the fault's location, severity, and energy. Low-energy events like partial discharge produce hydrogen, methane, and ethane. The sustained high-energy arcing results in the generation of all gases, with acetylene requiring the most energy. The appearance of these gases in transformers is influenced by temperature, as they dissolve within the insulation oil based on the nature and intensity of the failures (Syafuruddin and Nugroho, 2020).

TABLE 1 Transformer types of failure.

Types of failures	Abbreviations
No Failure	NM
Thermal Failure ($T < 300^\circ C$)	LO
Thermal Failure ($300^\circ C < T < 700^\circ C$)	MO
Thermal Failure ($T > 700^\circ C$)	HO
Discharges of Low Energy	LD
Discharges of High Energy	HD
Partial discharges	PD
Concurrent overheating and discharge	OD

An accurate fault diagnosis involves alerting to gases surpassing the standard limits in terms of concentrations, increments, rates of change, or ratios. The diagnostic report should include concise interpretive remarks and recommendations derived from these findings. To facilitate the identification of various faults during the diagnostic process, distinct classes have been established. These classes are determined through the physical examination of numerous faulty transformers, which are detectable through visual inspections and DGA outcomes. These classes are also summarised in Table 1.

Failures such as overheating, discharges, and Concurrent overheating and discharge generate various gases. The specific nature and intensity of the fault can be determined by analyzing the concentrations and types of these gases. The primary approach for diagnosing transformer failures using Dissolved Gas Analysis (DGA) involves extracting status information, which consists of various combinations of characteristic gases, throughout the operation of the power transformer. Subsequently, an existing diagnostic model algorithm is applied to analyze and differentiate this state information. Based on the judgment results, power system operators can discern the fault type and severity of the transformer. This information enables them to promptly devise specific countermeasures tailored to different situations, aiming to minimize fault-related losses within the power system to the greatest extent possible. The conventional methods include Doernenburg Ratio Method (CH_4/H_2 , C_2H_2/C_2H_4 , C_2H_2/CH_4 , C_2H_6/C_2H_2), Rogers Ratio Method (CH_4/H_2 , C_2H_2/C_2H_4 , C_2H_4/C_2H_6 , C_2H_6/CH_4), and Duval Triangle Method (CH_4/H_2 , C_2H_2/C_2H_4 , C_2H_4/C_2H_6).

2.1 Doernenburg Ratio Method

Doernenburg Method (DRM), which can be found in the IEEE C57.104–1991 guide, makes use of the ratios of the concentration of the key gases Hydrogen, Methane, Ethane, Ethylene, and Acetylene (Stenkovski et al., 2022). Although it was withdrawn by IEEE in 2006, it is one of the most effective methods in the diagnosis of electrical equipment such as transformers. This method is restricted by a few criteria. This diagnostic approach often has the drawback

TABLE 2 The limited concentration of the dissolved gas (Jongvilaikasem et al., 2022).

Types of gas	L1 concentration (ppm)
H ₂	100
CH ₄	120
CO	350
C ₂ H ₂	35
C ₂ H ₄	50
C ₂ H ₆	65

of leading to ‘no diagnosis’, a situation that happens more often than not (Jongvilaikasem et al., 2022). To apply this method, one of the key gases (H₂, C₂H₂, C₂H₄, C₂H₆, or CH₄) must have a concentration at least double the L1 concentration levels listed in Table 2. By referring to Table 3, the failure can be diagnosed.

2.2 Rogers Ratio Method

The Roger Ratio Method (RRM) utilizes gas ratios similar to those in the DRM, specifically C₂H₂/C₂H₄ and CH₄/H₂, with the exception of the ratio of C₂H₄/C₂H₆ (Bakar et al., 2014), as shown in Table 4. Unlike the DRM, the RRM can be used even if the key gases are not at high concentrations. This means that, with RRM, the gas concentrations do not need to be at least twice the L1 levels (Rogers, 1978). DRM and RRM are effective in diagnosing failures, but they have a limitation. Sometimes, certain gas combinations do not fit within the specified value range. When this happens, it becomes impossible to determine the type of fault.

2.3 Duval Triangle Method (DTM)

DTM works on the idea that various faults in a transformer create different gases. By measuring the amounts of these gases, it is possible to figure out what kind of fault is present (Duval, 2008). It came from IEC TC10 databases and an existing IEC 60599 Ratio method. The results of the gas analysis are then plotted on a triangular diagram, known as the Duval Triangle. This triangle is divided into different zones, each corresponding to a particular type of fault, such as thermal faults, electrical discharges, or a combination of both (Li and Zhang, 2016). The application of the DTM relies on three essential gases (CH₄, C₂H₄, and C₂H₂), which represent progressively higher energy levels of gas generation. The concentrations of these gases are computed and subsequently graphed on the three edges of a triangular diagram, as shown in Figure 1. Compared with DRM and RRM, DTM will not appear in cases where some results can fall outside the codes and no diagnostics can be given. DTM has created a closed-loop so it can always consistently yield a diagnosis, and maintain a low rate of incorrect assessments.

2.4 Digital methods

Several approaches have been suggested to enhance the precision of ratio-based DGA diagnostic approaches. These can generally be divided into fuzzy logic-based methods, heuristic methods and multiple data-driven artificial intelligence (AI) technologies.

Fuzzy logic facilitates the expression of imprecise requirements, encompassing uncertain criteria or loosely defined constraints, allowing for reasoning in situations with uncertain specifications. Tomsovic et al. (1993) introduced a theoretical fuzzy information model and inference scheme, utilizing rule-based representation to systematically integrate diverse diagnostic methods. This approach effectively resolved the potential rule conflicts and ensured the generation of the most consistent conclusions. Its advantages included robustness to missing data, ease of expansion to new diagnostic methods, and analytical performance evaluation. However, functions in fuzzy logic systems need to be determined based on expert experience, and effective fuzzy logic systems require comprehensive knowledge from human experts. In addition, the limitation also occurred when the rules in the rule-based system could not be automatically adjusted through the self-learning process, requiring more interventions from human experts.

Most of the intelligent fault diagnosis methods in power transformers are based on a variety of data-driven machine-learning technologies. Saravanan et al. (2020) applied a Multilayer Artificial Neural Network (ANN) to classify the transformer failure with an accuracy of 76%. However, Katooli and Koochaki (2020) indicated that the ANNs utilized BPA as a learning algorithm that reduced their level of accuracy. Some works applied the machine learning approaches as attribute selectors and combined them with the Support Vector Machine (SVM) to improve precision and reliability. Moreover, other than the SVM, some works implemented the ANN as the failure classifier. For example, Li et al. (2016) proposed a power transformer failure diagnosis approach based on an SVM achieving a diagnosis accuracy of 87.18%. In this approach, GA selected the free control parameters of SVM including penalty parameter, sigma, and type of kernel function, and then SVM was for transformer failures classification. Moreover, Dwiputranto et al. (2021) proposed a method that combined GA and ANN to classify the type of transformer failure reliably based on DGA data and achieved an accuracy of 95%.

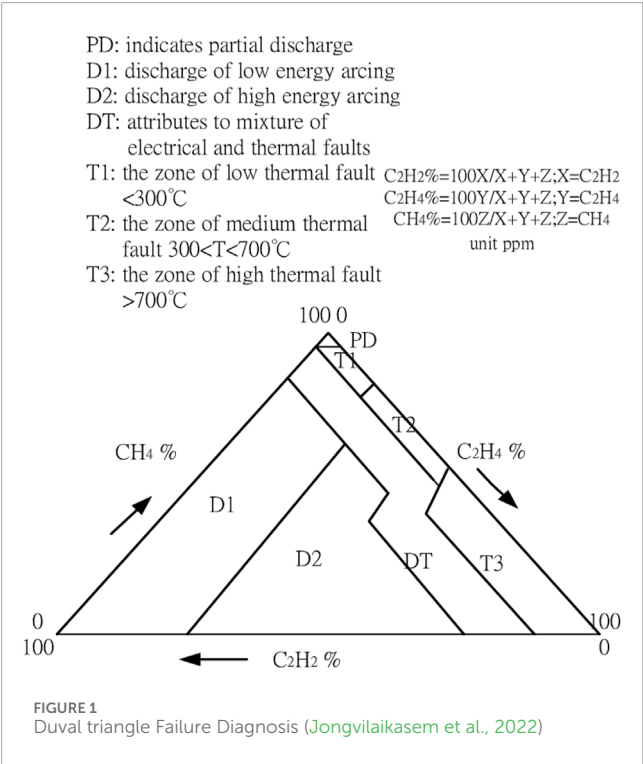
Numerous techniques have been introduced for diagnosing power transformer failures. However, most of the existing technologies utilise black-box models to obtain fault diagnosis results while the causality was relatively inconsiderate in this domain. When it comes to diagnosing faults and understanding the reasoning process, using black-box models often fails to provide explanations that match real-world phenomena. Explanations are rooted in the fundamental concept of causality and the absence of a causal framework in contemporary systems has been linked to issues related to generalization, interpretability, and outcome quantification (Peters et al., 2017). The exploration of techniques to distil rules from equipment failure events remained a developing area. This approach often lacked a deep understanding of the basic causes and struggled with diagnosing multiple failures and updating with new knowledge (Saravanan et al., 2020). Our approach aims to identify and understand how different types of failures cause various

TABLE 3 Doernenburg RATIO FAILURE DIAGNOSIS METHOD (Jongvilaikasem et al., 2022).

Potential failure	C_2H_4/H_2	C_2H_2/C_2H_4	C_2H_2/CH_4	C_2H_6/C_2H_2
Thermal failure	>1.0	<0.75	<0.3	>0.4
Partial discharge (low intensity)	<0.1	Not significant	<0.3	>0.4
high intensity discharge	0.1 to 1.0	>0.75	>0.3	<0.4

TABLE 4 Rogers ratio failure diagnosis method (Jongvilaikasem et al., 2022).

Potential failure	C_2H_2/C_2H_4	CH_4/H_2	C_2H_4/C_2H_6
Normal	< 0.1	0.1 to 1.0	<1.0
low intensity discharge	< 0.1	<0.1	<1.0
high intensity discharge	0.1 to 0.3	0.1 to 1.0	1.0 to 3.0
medium temperature thermal failure	< 0.1	>1.0	1.0 to 3.0
high temperature thermal failure	< 0.1	>1.0	3.0



gases to be produced. This helps in clearly determining the source and type of the failure.

3 Proposed method for transformer failure diagnosis

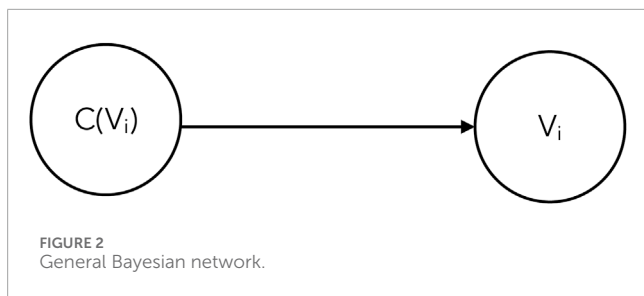
To solve the challenges of lacking transparency and training data, this research proposes a causal-based method to diagnose the

type of failure by discovering the causal relationship between the concentration and the types of the specific gases and the types of failures. To process the methodology, it would be necessary to review the causal reasoning first.

3.1 Causal reasoning

Causal reasoning, recognized as a fundamental cognitive capability, empowers individuals to navigate the complexities of the world by discerning the underlying causes of observed phenomena and devising effective problem-solving strategies (Waldmann, 2017). The objective of causal reasoning is to acquire adequate knowledge about the causal relationships between variables to predict/infer the outcome accurately (Stuart, 2010). Within this context, causal reasoning serves as a framework for systematically formalizing our comprehension of the data-generating process through the application of Structural Causal Models (SCMs). SCMs, or Structural Causal Models, help estimate how interventions affect the data produced by a certain process (Nogueira et al., 2022). To analyze the causal impacts of certain variables on others, a formalization of causal relationships is required. A widely accepted method for representing these relationships is through the utilization of a causal-directed acyclic graph (causal DAG), often referred to as a causal diagram (Shen et al., 2020). Such a diagram can capture prior assumptions regarding the causal structure of interest, which can include insights from expert knowledge. Usually, conventional causal reasoning focuses on figuring out the effects of certain causes. But this diagnostic system works the other way around: it starts from the effects and tries to find out what the possible causes might be.

Bayesian Networks (BNs), probabilistic graphical models that depict the probabilistic associations among random variables, are regarded as a potent instrument in the domain of causal reasoning (Lu et al., 2023). Moreover, Lakehal et al. (2015) illustrates that BNs



allow for the reversal of probabilities, enabling the identification of causes when effects observations and cause consequences are known. Refer to [Figure 2](#), a Bayesian Network contains an edge connecting two variables signifies a direct dependence between them, with one acting as the parent (cause) and the other as the child (effect). To capture this relationship, each node in the network possesses a conditional probability table, quantifying the child variable's behavior in response to its parent(s) if multiple parents are involved. However, for root nodes without parents, the probability table is non-conditional, representing prior probabilities related to variable values. The graphical representation is referred to as the *model structure*, with associated probability tables termed *model parameter*. Therefore, identifying the cause-and-effect relationships and conducting the causal model requires both structural learning and parameter learning. These structural elements can either be derived from expert knowledge or computed based on available data.

BNs effectively handle uncertainty and depict interrelationships between problem variables, offering a visually accessible representation and enabling comprehensive failure probability analysis for complex systems ([Lakehal et al., 2015](#); [Özgür-Ünlüakın et al., 2021](#)). Consequently, BNs serve as the foundational framework underpinning this paper. By leveraging BNs, our approach discovers the relationship between the types and quantities of specific gases under various power transformer states. Integrating this relationship under the dissolved gas analysis framework, the invention establishes the BNs model for power transformer state diagnosis. Our proposed diagnosis approach will be introduced in detail in [Section 3.2](#).

3.2 Methodology

Our proposed method, integrating causal reasoning and dissolved gas analysis. It aims to work well in general, using a small amount of data to evaluate the condition and track faults in various types of power transformers. As shown in [Figure 3](#). The proposed method comprises three main steps: 1) Conducting a Causal model; 2) Training a Bayesian Network with structure learning and parameter learning; and 3) Utilizing the model to diagnose failures. For this study, a quantity of 823 samples was obtained from the [North China Electric Power Research Institute \(2021\)](#), involving the transformers in 110 kV, 220 kV, 330 kV, 500 kV, and 750 kV. This dataset consists of a subset of 727 samples for the model training and a subset of 96 samples for the testing. After the conduction of the model, the proposed method will be tested

through the testing dataset with a comparison with the conventional methods, including DRM, RRM, and DTM. The following will introduce the process of model conduction.

3.2.1 Causal model conduction

Given the differences in initial data from various transformers, it is crucial to preprocess this data. This step aligns the inputs with the model, minimizing noise and ensuring the accuracy of the causal model. Addressing the issue of limited failure data and the low interpretability of power grid transformers, our method is proposed to refine by discovering and constructing causal relationships between transformer status, gas production types, and quantities. The BNs serve as the foundational method for fault diagnosis and traceability. In light of inevitable errors from causal learning, manual intervention, branch reduction, and optimization through expert experience are implemented to enhance model accuracy. Ultimately, a fault reasoning method is formulated based on the cause-and-effect model of power transformer status and gas production. This method conducts explainable reasoning, failure-type judgment, and screening to support operational, maintenance, and repair activities. This enhancement seeks to enhance work quality and efficiency in power transformers.

[Diehl and Ramirez-Amaro \(2022\)](#) indicate that the process of explaining failures necessitates the acquisition of knowledge regarding the cause-and-effect relationship between potential causes and the effects of a phenomenon. The framework of this model is established based on the DGA technology. Therefore, the potential causes should be various transformer states (refer to [Table 1](#)) and the effects should be the type and the content of the dissolved gases in transformer oil. According to ([Feng et al., 2021](#)), due to the rareness of safety-critical events, adversarial learning proves valuable in significantly reducing the necessary data quantity without compromising impartial evaluation. In other words, an appropriate reduction of the normal data subset can substantially reduce the overall dataset size without reducing the capacity of the model. Therefore, the training dataset was designed as shown in [Table 5](#).

Moreover, in the analysis of gas release, seven types of dissolved gases are commonly considered: H_2 , CH_4 , C_2H_2 , C_2H_4 , C_2H_6 , CO , and CO_2 ([Gouda et al., 2016](#)). However, to reduce dataset requirements and simplify the model, the investigation of the two carbon oxides (CO , CO_2) will be eliminated as their relatively minor proportional variations.

This research includes multiple transformers, and gas production levels may vary. To address this, data preprocessing will use gas proportions instead of the content. Unlike traditional ratio-based DGA or other methods that use normalization, our method calculates specific gas ratios to the total gas, offering comprehensive observations. A sudden increase in a specific gas raises its proportion, reducing the ratios of other gases. H_2 , with significant overall presence, is considered separately. Proportions are calculated as follows:

$$H_2\% = \frac{H_2}{H_2 + CH_4 + C_2H_2 + C_2H_4 + C_2H_6} \quad (1)$$

$$CH_4\% = \frac{CH_4}{CH_4 + C_2H_2 + C_2H_4 + C_2H_6} \quad (2)$$

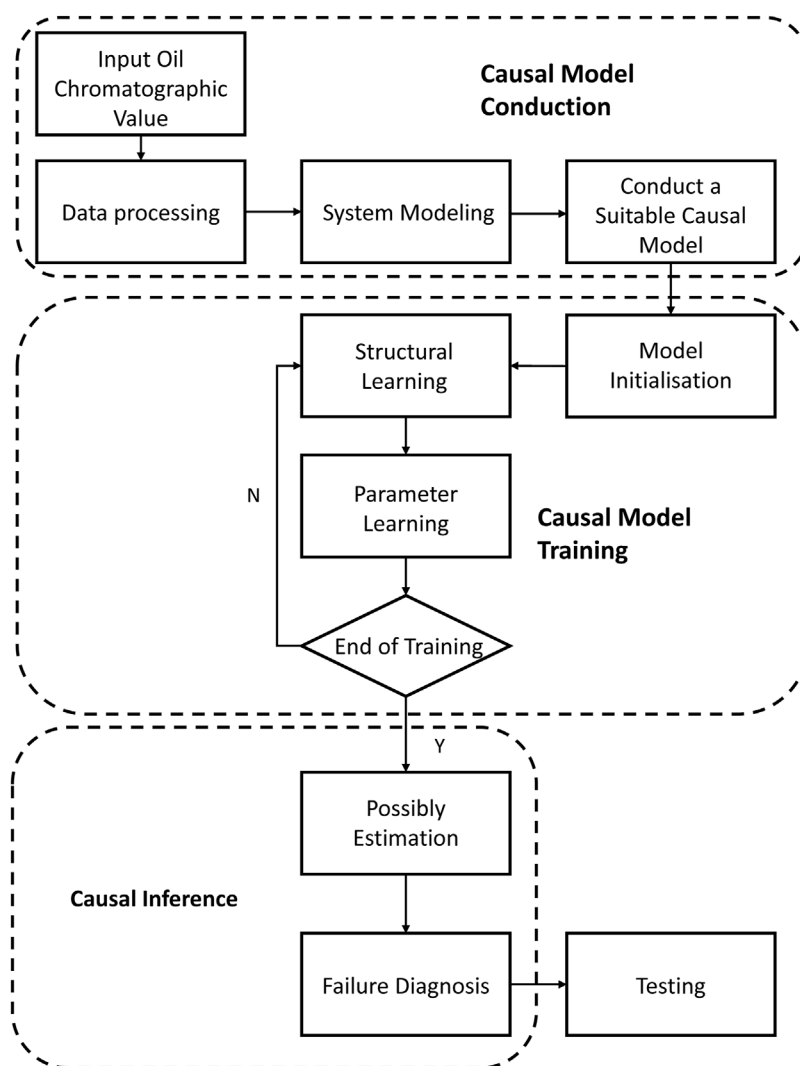


FIGURE 3
General causal model for transformer failure diagnosis.

$$C_2H_2\% = \frac{C_2H_2}{CH_4 + C_2H_2 + C_2H_4 + C_2H_6} \quad (3)$$

$$C_2H_4\% = \frac{CH_4}{CH_4 + C_2H_2 + C_2H_4 + C_2H_6} \quad (4)$$

$$C_2H_6\% = \frac{CH_4}{CH_4 + C_2H_2 + C_2H_4 + C_2H_6} \quad (5)$$

After identifying the variables, the next step is to construct the causal model. A set of treatments X will be generated, and denoted as $C \subset X$, representing potential causes, along with outcome variables identified as $E \subset X$. The primary objective of causal inference will be to assess and quantify the impact of C on E and utilize the impact to trace back the potential states. To concise the training process, our method absorbs the benefits from fuzzy logic concepts and processes the computed proportion data of the gases (the outcomes of Eq. 1-5) into six intervals, denoted as X_1 to X_6 , effectively representing severity. Consequently, the input of the model will,

therefore, become the interval index. Its primary objective is to learn how the gas release severity reflects the current state type.

3.3 Model training

As mentioned previously, this study adopts BNs as the framework of the causal model conduction, due to the benefits such as their capability for the reversal of inference. BNs are characterized by a graphical structure denoted as $G=(V, A)$, which forms a directed acyclic graph (DAG). In which, $V = \{X_1, X_2, \dots, X_n\}$, refer to the set of nodes and A refer to the set of arcs. By utilizing the dependency structure in the Directed Acyclic Graph (DAG) and the Markov property, the joint probability distribution of a Bayesian Network (BN) can be broken down into a set of local probability distributions. In this situation, each random variable X_i on its direct parents ΠX :

$$p(X_1, X_2, \dots, X_n) = \prod p(X_i | \text{parents}(X_i)) \quad (6)$$

TABLE 5 Transformer STATE types in the training samples.

State types (abbr.)	Number of samples
NM	127
LO	19
MO	39
HO	88
LD	70
HD	298
PD	25
OD	61

Learning the Bayesian network from the data requires two steps: structure learning and parameter learning (Diehl et al., 2021). Structure learning is dedicated to learning the graphical structure $G=(V, A)$. Refer to Figure 4, learning the structure of the BNs is to identify the different types of failures associated with the proportion of the target gases. There are two families of the most common methods to learn the graphical model including constraint-based algorithms and score-based algorithms. Constraint-based algorithms consider conditional independence constraints, whereas score-based algorithms generate multiple potential causal graphs, assign scores to each, and ultimately choose a final graph based on these scores (Shen et al., 2020). In this research, the score-based algorithm was chosen because, despite its higher computational time demands, it has been demonstrated to exhibit greater accuracy in learning causal structures (Shen et al., 2020). It is important to highlight that acquiring reliable assumptions regarding causal relationships represents a significant challenge within the broader domain of causal inference (Sharma et al., 2021). The gold standard causal structure for the transformer system was not available. As a result, the resulting causal structure will be presumed correct or manually adjusted based on domain knowledge.

Ji et al. (2015) indicate that structural learning is about the exploration of casual relationships among a multitude of variables, and parametric learning focuses on the estimation of conditional probability tables, which capture probabilistic relationships between variables. Moreover, the computation of the desired probabilities within the model becomes feasible only once both the structure and parameters have been defined. In this study, the maximum likelihood estimator will be employed to construct a conditional probability table, leveraging the previously established network structure.

3.4 Utilizing model to diagnose failures

In this research, the causal model utilizes the BNs to compute the possibility of the occurrence of the states. By training the historical data of the transformers, the model will be able to determine the occurrence probabilities of each state, based on the

conditional probability tables. The probability of occurrence is computed from the formula (6) ultimately enabling the diagnosis of transformer failures.

In this approach, BNs were employed as the causal model to calculate the occurrence probability of the states. By inputting and processing gas production data from power transformers, the model determines the probability of each state based on conditional probability tables. The probability of occurrence is computed according to the formula (6), enabling the diagnosis of transformer faults. To address data limitations, the model simplifies training by streamlining the inference process. In Figure 3, fault types are initially classified through expert knowledge into fault type I and fault type II, addressing diagnostic challenges and enhancing accuracy. The proposed method enables multi-layer fault diagnosis and traceability for current power transformers, identifying fault types through intelligent and interpretable reasoning based on known gas production conditions.

4 Model testing and discussion

The testing revolves around the utilization of 96 authentic samples, which involve the transformers in 110 kV, 220 kV, 330 kV, 500 kV, and 750 kV. Each of these testing samples has been painstakingly labelled with its corresponding failure type. The primary objective of this testing is to examine our diagnosis model by analyzing these data and then comparing its diagnostic results with the existing labelled state types. The following presents an example of the testing of the proposed model.

4.1 Example presentation

Taking an example using the data in Table 6, the system follows the evaluation procedure (as shown in Figure 4) during the diagnostic process.

The probabilities of occurrence for each STATE TYPE II are illustrated in Table 7.

It can be seen that Discharge presents the highest possibility of occurrence. Therefore, the system will then move to STATE TYPE II, and diagnose the failure among HD, LD, and PD. The probabilities of occurrence for each STATE TYPE I are illustrated in Table 8.

After the BN inference, posterior probabilities given by the network are presented in TABLE8. In this table, the posterior probability of the high_energy discharge is 0.78, a higher probability compared to the probability of low_energy discharge (0.13) and partial discharge (0.09). Based on this inference, the model is possible to make a conclusive decision, confirming the state type as high_energy discharge. The actual description of the state is a high_energy discharge caused by the solder joints of the potential connection wire falling off, and the diagnosis results are consistent with it.

4.2 Testing

The objective of the testing is to evaluate the diagnostic accuracy of the system in identifying and classifying transformer failures. In

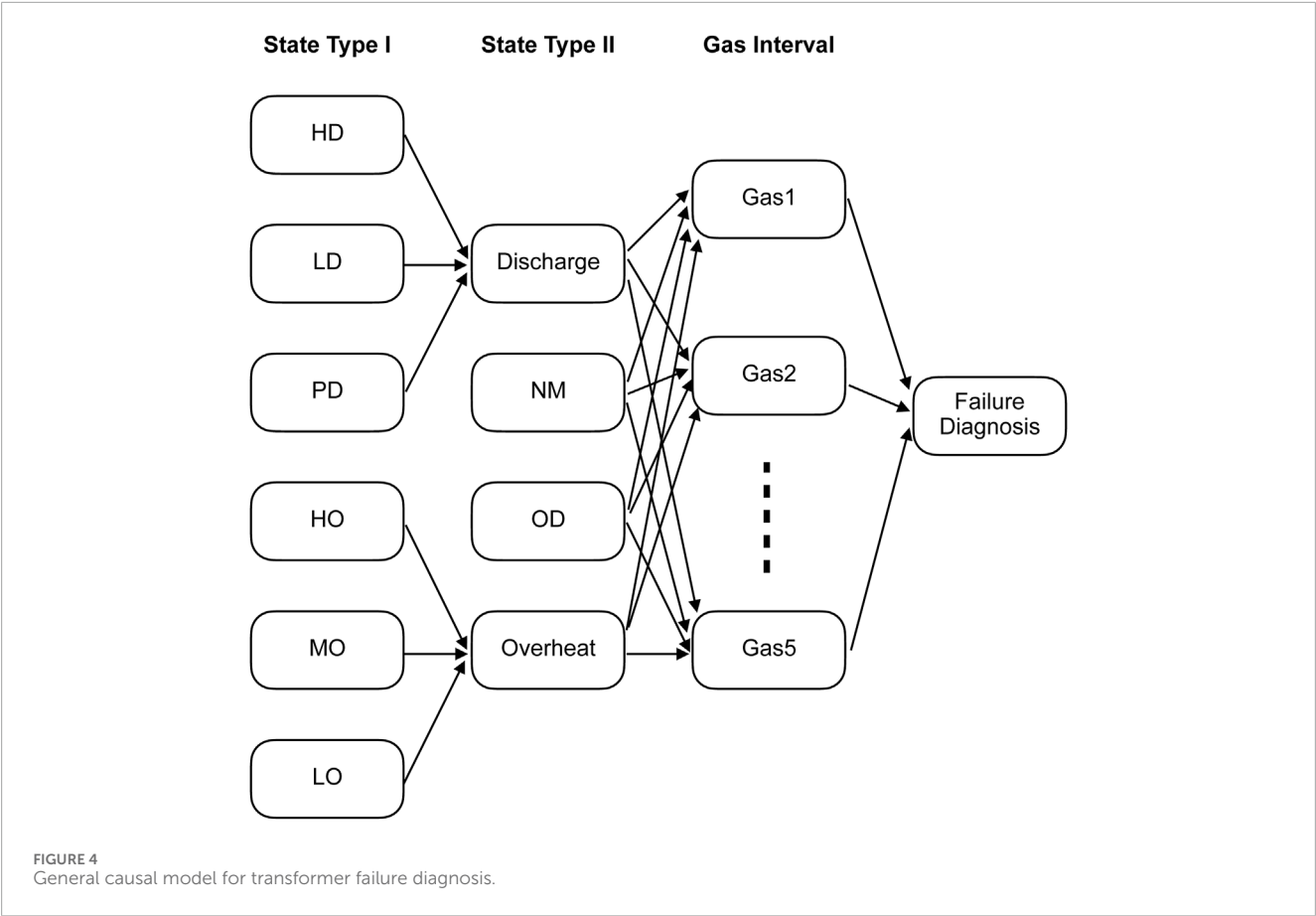


TABLE 6 An example of the testing dataset.

H ₂	CH ₄	C ₂ H ₂	C ₂ H ₄	C ₂ H ₆	Failure type
443.99	116.38	5.33	7.52	91.03	High_energy Discharge

TABLE 7 List of possibilities for each State Type II.

STATE TYPE II	Possibility
NM	0.08
Overheating	0.05
Discharge	0.70
OD	0.17

TABLE 8 List of possibilities for each State Type I.

STATE TYPE I	Possibility
LD	0.13
HD	0.78
PD	0.09

the assessment of the testing, four systems were tested by analyzing 96 individual samples. The results are listed in Table 9. The Equation of accuracy and % unsolved is as follows:

$$Accuracy = \frac{\text{number of correct diagnosis}}{\text{number of diagnosis}} \times 100\% \quad (7)$$

$$\%Unsolved = \frac{\text{number of unsolved diagnosis}}{\text{number of diagnosis}} \times 100\% \quad (8)$$

The evaluation outcomes presented in Table 9 indicate that among the evaluated methods, the DTM achieves the highest level of accuracy. While DRM, RRM, and DTM all demonstrate commendable accuracy, the %Unsolved indicates that they encounter challenges in diagnosing certain scenarios, particularly NM conditions. For instances, these methods can not identify an ‘NM’ state, consistently indicating the presence of a fault instead. Moreover, when certain gas combinations fall outside the designated value range upon calculation, DRM and RRM was impossible to identify the fault type. Consequently, their diagnosis presents a sector of unsolved.

The proposed causal system attained an impressive accuracy rate of 92 out of 96 cases. It not only showed good accuracy compared to conventional methods but also demonstrated the capability to discern system failures. Moreover, both RRM and DRM present less accuracy in diagnosing the LO condition while DTM and Causal method performs better. It is important to acknowledge that variations in the testing dataset

TABLE 9 Number of correct diagnosis among different methods.

Method	Number of diagnosis	Number of correct diagnosis	Accuracy (%)	% unsolved
DRM	60	58	97.1	37.5
RRM	55	51	92.7	42.7
DTM	81	80	98.8	15.6
Causal	96	92	95.8	0

contributed to differences in accuracy. Despite this, the results clearly show that the model accurately identified the state types in almost all tested instances across various transformer types, underscoring its practical effectiveness. Besides, unlike the one-size-fits-all approach of conventional methods, the proposed method presents the probabilities of occurrence of all states and provides higher flexibility for more personalized and accurate assessments.

To explore the failures, notably, in some instances, failures occur when the actual failure type is the second most likely failure type predicted by the system. Moreover, failures also happen during the distinction between HO and MO. One possible reason might be the challenge due to the variability in the data. Temperature measurements in real-world transformer operations can fluctuate, making it difficult to establish clear thresholds for categorization. The employment of interval data preprocessing would also affect the accuracy of the system, the exact distinction of each interval should be tuned manually. This highlights a potential area for further improvement or manual involvement in the diagnostic process. Besides, these situations highlight the advantage of utilizing BNs in this diagnostic context, which is the transparency it offers across all possible scenarios. This transparency can be instrumental in understanding the system's reasoning and decision-making process, enabling engineers and experts to analyze and intervene when necessary, further enhancing the system's diagnostic capabilities. Besides, the proposed Bayesian model offers the added benefit of simultaneous diagnosis of two failures, for example, the diagnosis of OD.

5 Conclusion

This paper proposes an approach to transformer failure diagnosis based on the relationship between dissolved gases in transformer oil and the type of failures in electric power operations. While previous studies primarily focused on failure classification, this study proposes an approach integrating causality and the DGA framework to infer power transformer failures. The proposed causal reasoning method can diagnose all potential transformer states, accounting for varying degrees of severity in both electrical and thermal failures. By comparing with the conventional ratio-based DGA methods including DTM, DRM, and RRM, the system was tested with a good accuracy of 95.8%. Furthermore, our model demonstrates proficiency in diagnosing various types of failures from different types of transformers, offering robust failure detection and classification, and the ability to diagnose multiple failures simultaneously. Besides, this study also extends the applicability

of causal reasoning in the domain of power systems. Future work will explore the failure prediction by employing the time-series data.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

FJ: Writing—original draft, Project Administration. ZM: Writing—review and editing, Data curation and Resources. QC: Conceptualisation, Methodology, Data curation, Writing—original draft. FZ: Funding acquisition, Writing—review and editing. DZ: Writing—original draft, Writing—review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Science and Technology Program of the Headquarters of State Grid Corporation of China, Research on Causal Reasoning Technology Based on the Electric Power Equipment Knowledge Graph, under Grant 5700-202255 291A-2-0-QZ. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aizpurua, J. I., Catterson, V. M., Stewart, B. G., McArthur, S. D., Lambert, B., Ampofo, B., et al. (2018). Power transformer dissolved gas analysis through bayesian networks and hypothesis testing. *IEEE Trans. Dielectr. Electr. Insulation* 25, 494–506. doi:10.1109/tdei.2018.006766
- Ali, M. S., Bakar, A. H. A., Omar, A., Jaafar, A. S. A., and Mohamed, S. H. (2023). Conventional methods of dissolved gas analysis using oil-immersed power transformer for fault diagnosis: a review. *Electr. Power Syst. Res.* 216, 109064. doi:10.1016/j.epsr.2022.109064
- Bakar, N. A., Abu-Siada, A., and Islam, S. (2014). A review of dissolved gas analysis measurement and interpretation techniques. *IEEE Electr. Insul. Mag.* 30, 39–49. doi:10.1109/mei.2014.6804740
- Diehl, M., Paxton, C., and Ramirez-Amaro, K. (2021). “Automated generation of robotic planning domains from observations,” in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 6732–6738.
- Diehl, M., and Ramirez-Amaro, K. (2022). Why did i fail? a causal-based method to find explanations for robot failures. *IEEE Robotics Automation Lett.* 7, 8925–8932. doi:10.1109/lra.2022.3188889
- Duval, M. (2008). The duval triangle for load tap changers, non-mineral oils and low temperature faults in transformers. *IEEE Electr. Insul. Mag.* 24, 22–29. doi:10.1109/mei.2008.4665347
- Dwiputranto, T. H., Setiawan, N. A., and Adji, T. B. (2021). “Dga-based early transformer fault detection using ga-optimized ann,” in 2021 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP) (IEEE), 342–347.
- Feng, S., Yan, X., Sun, H., Feng, Y., and Liu, H. X. (2021). Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment. *Nat. Commun.* 12, 748. doi:10.1038/s41467-021-21007-8
- Gouda, O. E., Salem, S., and El-Hoshi, S. H. (2016). Power transformer incipient faults diagnosis based on dissolved gas analysis. *TELKOMNIKA Indones. J. Electr. Eng.* 17, 10–16. doi:10.11591/ijeecs.v16.i3
- Guo, C., Dong, M., and Wu, Z. (2019). “Fault diagnosis of power transformers based on comprehensive machine learning of dissolved gas analysis,” in 2019 IEEE 20th International Conference on Dielectric Liquids (ICDL) (IEEE), 1–4.
- Ji, Z., Xia, Q., and Meng, G. (2015). “A review of parameter learning methods in bayesian network,” in Advanced Intelligent Computing Theories and Applications: 11th International Conference, ICIC 2015 Proceedings, Part III 11, Fuzhou, China, August 20–23, 2015 (Springer), 3–12.
- Jongvilaikasem, K., Pattanadech, N., Wattakapaiboon, W., Kando, M., Maneerot, S., and Pannil, P. (2022). The comparison of dga interpretation techniques application for actual failure transformer inspections including experience from power plants in Thailand. *Int. J. Electr. Eng. Inf.* 14, 224–233. doi:10.15676/ijeeci.2022.14.1.14
- Kari, T., Gao, W., Zhao, D., Zhang, Z., Mo, W., Wang, Y., et al. (2018). An integrated method of anfis and dempster-shafer theory for fault diagnosis of power transformer. *IEEE Trans. Dielectr. Electr. Insulation* 25, 360–371. doi:10.1109/tdei.2018.006746
- Katooli, M. S., and Koochaki, A. (2020). Detection and classification of incipient faults in three-phase power transformer using dga information and rule-based machine learning method. *J. Control, Automation Electr. Syst.* 31, 1251–1266. doi:10.1007/s40313-020-00625-5
- Lakehal, A., Ghemari, Z., and Saad, S. (2015). “Transformer fault diagnosis using dissolved gas analysis technology and bayesian networks,” in 2015 4th International Conference on Systems and Control (ICSC) (IEEE), 194–198.
- Lewis, D. (1986). “Causal explanation,” in *Philosophical papers vol. 2* (Oxford University Press), 214–240.
- Li, J., Zhang, Q., Wang, K., Wang, J., Zhou, T., and Zhang, Y. (2016). Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine. *IEEE Trans. Dielectr. Electr. Insulation* 23, 1198–1206. doi:10.1109/tdei.2015.005277
- Li, Y., and Zhang, G.-J. (2016). “Combining conventional and artificial intelligence dga interpretation methods using optimized weighting factor,” in 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA) (IEEE), 37–42.
- Lu, Y., Zheng, Q., and Quinn, D. (2023). Introducing causal inference using bayesian networks and do-calculus. *J. Statistics Data Sci. Educ.* 31, 3–17. doi:10.1080/26939169.2022.2128118
- Min, L. H., and Chang, C. (2009). “Application of dempster-shafer’s theory of evidence for transformer incipient fault diagnosis,” in 8th International Conference on Advances in Power System Control, Operation and Management (APSCOM 2009), Hong Kong, China, 08–11 November 2009 (IET). doi:10.1049/cp.2009.1844
- Nogueira, A. R., Pugnana, A., Ruggieri, S., Pedreschi, D., and Gama, J. (2022). Methods and tools for causal discovery and causal inference. *Wiley Interdiscip. Rev. data Min. Knowl. Discov.* 12, e1449. doi:10.1002/widm.1449
- North China Electric Power Research Institute (2021). *Transformer oil chromatographic analysis and fault diagnosis case*. Beijing, China: China Electric Power Press.
- Özgür-Ünlüakın, D., Türkali, B., and Aksezer, S. Ç. (2021). Cost-effective fault diagnosis of a multi-component dynamic system under corrective maintenance. *Appl. Soft Comput.* 102, 107092. doi:10.1016/j.asoc.2021.107092
- Pearl, J., and Mackenzie, D. (2018) *The book of why: the new science of cause and effect*. Basic Books.
- Peters, J., Janzing, D., and Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms*. The MIT Press.
- Rogers, R. (1978). Ieee and iec codes to interpret incipient faults in transformers, using gas in oil analysis. *IEEE Trans. Electr. insulation* 13, 349–354. doi:10.1109/TEL.1978.298141
- Saravanan, D., Hasan, A., Singh, A., Mansoor, H., and Shaw, R. N. (2020). “Fault prediction of transformer using machine learning and dga,” in 2020 IEEE international conference on computing, power and communication technologies (GUCON) (IEEE), 1–5.
- Sharma, A., Syrgkanis, V., Zhang, C., and Kiciman, E. (2021). Dowhy: addressing challenges in expressing and validating causal assumptions. arXiv preprint arXiv:2108.13518.
- Shen, X., Ma, S., Vemuri, P., Simon, G., Weiner, M. W., Aisen, P., et al. (2020). Challenges and opportunities with causal discovery algorithms: application to alzheimer’s pathophysiology. *Sci. Rep.* 10, 2975. doi:10.1038/s41598-020-59669-x
- Stenkovski, T., Mojsoska, N., and Arapinoski, B. (2022). *Methods of analysis of dissolved gasses in transformer oil*, 1–4.
- Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Stat. Sci. a Rev. J. Inst. Math. Statistics* 25, 1–21. doi:10.1214/09-sts313
- Syafruddin, H., and Nugroho, H. P. (2020). “Dissolved gas analysis (dga) for diagnosis of fault in oil-immersed power transformers: a case study,” in 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia, 03–04 September 2020 (IEEE), 57–62.
- Tomsovic, K., Tapper, M., and Ingvarsson, T. (1993). A fuzzy information approach to integrating different transformer diagnostic methods. *IEEE Trans. Power Deliv.* 8, 1638–1646. doi:10.1109/61.252690
- Waldmann, M. (2017). *The Oxford handbook of causal reasoning*. Oxford University Press.



OPEN ACCESS

EDITED BY

Hengrui Ma,
Qinghai University, China

REVIEWED BY

Zulfiqar Ahmad Khan,
Sejong University, Republic of Korea
Yassine Himeur,
University of Dubai, United Arab Emirates
Wei Gao,
University of Denver, United States

*CORRESPONDENCE

Xiaoxu Hu,
✉ dearhxx1@126.com

RECEIVED 29 November 2023

ACCEPTED 07 March 2024

PUBLISHED 18 June 2024

CITATION

Liu M and Hu X (2024), LoadNet: enhancing energy storage system integration in power system operation using temporal convolutional and recurrent models with self-attention.
Front. Energy Res. 12:1346398.
doi: 10.3389/fenrg.2024.1346398

COPYRIGHT

© 2024 Liu and Hu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

LoadNet: enhancing energy storage system integration in power system operation using temporal convolutional and recurrent models with self-attention

Minggang Liu and Xiaoxu Hu*

Department of Computer Science, Harbin Finance University, Harbin, China

Introduction: In the context of the evolving energy landscape, the efficient integration of energy storage systems (ESS) has become essential for optimizing power system operation and accommodating renewable energy sources.

Methods: This study introduces LoadNet, an innovative approach that combines the fusion of Temporal Convolutional Network (TCN) and Gated Recurrent Unit (GRU) models, along with a self-attention mechanism, to address the challenges associated with ESS integration in power system operation. LoadNet aims to enhance the management and utilization of ESS by effectively capturing the complex temporal dependencies present in time-series data. The fusion architecture of TCN-GRU in LoadNet enables the modeling of both short-term and long-term dependencies, allowing for accurate representation of dynamic power system behaviors. Additionally, the incorporation of a self-attention mechanism enables LoadNet to focus on relevant information, facilitating informed decision-making for optimal ESS operation. To assess the efficacy of LoadNet, comprehensive experiments were conducted using real-world power system datasets.

Results and Discussion: The results demonstrate that LoadNet significantly improves the efficiency and reliability of power system operation with ESS. By effectively managing the integration of ESS, LoadNet enhances grid stability and reliability, while promoting the seamless integration of renewable energy sources. This contributes to the development of a more sustainable and resilient power system. The proposed LoadNet model represents a significant advancement in power system management. Its ability to optimize power system operation by integrating ESS using the TCN-GRU fusion and self-attention mechanism holds great promise for future power system planning and operation. Ultimately, LoadNet can pave the way for a more sustainable and efficient power grid, supporting the transition to a clean and renewable energy future.

KEYWORDS

multi-modal data, microgrid, knowledge-driven power scenario understanding technology, neural network fusion, virtual power plant

1 Introduction

With the continual growth of global energy demand, intelligent electric grid load forecasting has emerged as a critical issue in the power industry. Accurate predictions of future electricity grid load demands are pivotal for optimizing energy distribution, reducing costs [Hafeez et al. \(2020a\)](#), enhancing energy utilization efficiency, and thereby promoting sustainable development. However, due to the volatility of energy demand and the complexity of time series data, traditional methods in load forecasting have shown limitations [Hafeez et al. \(2021\)](#). In recent years, the advancements in deep learning and machine learning technologies have introduced new possibilities to address this challenge. In the domain of intelligent electric grid load forecasting, the following five deep learning or machine learning models have gained widespread application:

a. Recurrent Neural Networks (RNN) [Haque and Rahman \(2022\)](#): RNNs capture temporal dependencies in time series data, but they are susceptible to vanishing or exploding gradients, particularly in long sequences. b. Gated Recurrent Units (GRU) [Shi et al. \(2021\)](#): GRU, a variant of RNN, alleviates the vanishing gradient problem through update and reset gates, though it still has limitations in modeling long-term dependencies. c. Long Short-Term Memory Networks (LSTM) [Li et al. \(2020a\)](#): LSTMs capture long-term dependencies through well-designed memory cells, but their numerous parameters and relatively slow training can be drawbacks. d. Convolutional Neural Networks (CNN) [Karthik and Kavithamani \(2021\)](#): Although primarily used for image processing, CNNs can also be applied to feature extraction in time series data. However, they may not effectively handle temporal relationships. e. Self-Attention Mechanism (Transformer) [Wang et al. \(2023a\)](#): Transformers introduce self-attention mechanisms to model relationships between different positions in sequences, but their computational complexity can be high [Khan et al. \(2023\)](#).

Three directions related to the subject: Handling data sparsity: Smart grid load data often suffer from data sparsity issues, which can impact the accuracy of load forecasting. Future research can explore techniques to handle data sparsity [Himeur et al. \(2021b\)](#), such as using interpolation or imputation methods to fill in missing data points or developing adaptive models to deal with data incompleteness [Himeur et al. \(2020\)](#), thereby improving the accuracy of load forecasting. Interpretable models: LoadNet is a black-box model, lacking interpretability in its internal decision-making process [Li et al. \(2023\)](#). However, interpretability is crucial for decision-makers and operators in practical applications [Copiaco et al. \(2023\)](#). Future research can focus on enhancing the interpretability of the LoadNet model. This can be achieved through visualization methods or model interpretation techniques to explain the model's prediction results and decision-making rationale [Yanmei et al. \(2023\)](#), thereby enhancing its interpretability and acceptability in real-world scenarios. Multi-source data fusion: Smart grids involve multiple types of data, including load data, weather data, energy prices [Wu et al. \(2022\)](#). Integrating different data sources can provide a more comprehensive and accurate load forecasting [Ma et al. \(2023\)](#). Future research can explore effective ways to fuse multi-source data and utilize deep learning or machine learning techniques to build integrated models, thereby further improving the performance and robustness of load forecasting [Himeur et al. \(2021a\)](#). Further research in these directions will

contribute to the advancement of load forecasting in smart grids, providing more accurate, reliable, and interpretable methods for load prediction.

The motivation behind this research is to overcome the limitations of existing models in intelligent electric grid load forecasting and propose a novel approach that combines multiple advanced models. Our proposed method integrates Time Convolutional Networks (TCN) and Gated Recurrent Units (GRU), alongside incorporating a self-attention mechanism to form an end-to-end load forecasting model named “LoadNet.” Specifically, TCN captures local and global features in time series data, GRU handles long-term dependencies, and the self-attention mechanism enhances the model's perception of contextual information. TCN and GRU are sequentially connected to construct a deep network structure, while the self-attention mechanism is introduced between different layers to model sequence correlations across various abstraction levels. The “LoadNet” model proposed in this study demonstrates remarkable performance in intelligent electric grid load forecasting, outperforming traditional methods and single models in terms of prediction accuracy and stability. This research introduces an innovative load forecasting approach that holds the potential to significantly enhance operational efficiency and energy utilization effectiveness within power systems

- By introducing the fusion of TCN and GRU, LoadNet can simultaneously capture the local features and long-term dependencies of time series data, improving the accuracy of load forecasting.
- The introduction of the self-attention mechanism helps to learn the relationship and importance between different time steps in the sequence, further improving the performance of the LoadNet model.
- Through experimental verification, LoadNet has achieved significant improvement on real smart grid load datasets, proving its potential and effectiveness in practical applications.

2 Methodology

2.1 Overview of our network

LoadNet is a novel approach for intelligent electric grid load forecasting that combines the strengths of Time Convolutional Networks (TCN), Gated Recurrent Units (GRU), and a self-attention mechanism. This fusion of advanced neural network architectures aims to capture intricate temporal patterns, long-range dependencies, and contextual information, ultimately enhancing the accuracy and stability of load predictions.

[Figure 1](#) shows the overall framework of our proposed method. Detailed Method Implementation:

- Input Data Preparation:

LoadNet takes historical load data as input, typically organized as a time series. The dataset is divided into training, validation, and test sets.

- Time Convolutional Networks (TCN):

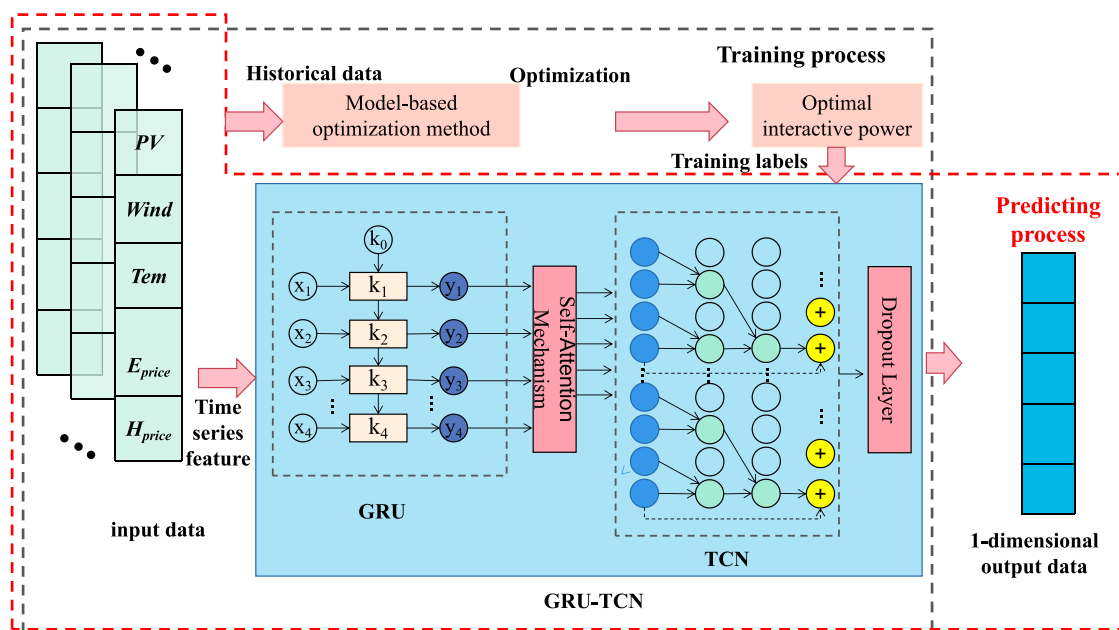


FIGURE 1
The overall framework of our proposed method.

TCN is employed as the initial feature extractor. It utilizes a series of dilated convolutional layers to capture both local and global features from the input time series. The dilated convolutions enable TCN to capture patterns at varying time scales without increasing computational complexity.

- Gated Recurrent Units (GRU):

To address long-term dependencies, GRU is integrated after TCN. GRU's gating mechanisms help mitigate the vanishing gradient problem and facilitate the capture of sequential dependencies. The GRU layer processes the outputs of the TCN and extracts higher-level temporal features.

- Self-Attention Mechanism:

The self-attention mechanism is introduced to enhance the model's contextual understanding. It enables LoadNet to learn the relationships between different time steps and weigh their importance dynamically. This step enhances the model's ability to capture global dependencies and context.

- Model Fusion and Hierarchical Representation:

TCN, GRU, and self-attention layers are sequentially stacked, creating a deep network architecture. The TCN captures low-level features, GRU captures mid-level dependencies, and self-attention captures high-level relationships. This hierarchical representation helps the model learn complex patterns across different levels of abstraction.

- Loss Function and Training:

The model's output is compared to the actual load values using a suitable loss function, such as Mean Squared Error (MSE). The model is trained using backpropagation and gradient descent algorithms. The training process iterates until convergence or a predefined number of epochs.

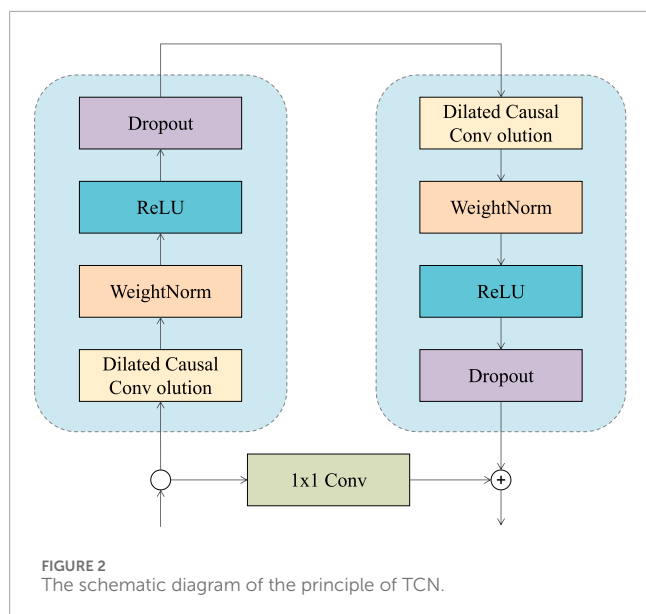
- Prediction and Evaluation:

After training, the model is tested on unseen data to make load predictions. The performance is evaluated using metrics like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

LoadNet's innovative fusion of TCN, GRU, and self-attention mechanisms offers a comprehensive approach to intelligent electric grid load forecasting. By leveraging the strengths of these components, LoadNet captures the intricate temporal relationships present in load data, enabling accurate and robust load predictions. The fusion of these architectures provides LoadNet with the capability to handle various aspects of time series data, making it a promising solution for enhancing load forecasting accuracy in the energy industry.

2.2 TCN network

Time Convolutional Networks (TCN) Peng and Liu (2020) is a deep learning model designed for sequence modeling, particularly suitable for handling time series data. The core idea behind TCN



is to capture patterns and features within time series by stacking multiple layers of one-dimensional convolutions. Unlike traditional recursive structures [Zhou et al. \(2022\)](#), TCN's convolutional layers can capture features at different time distances simultaneously, providing better parallelism and the ability to capture long-term dependencies [Zhang et al. \(2023\)](#). [Figure 2](#) is a schematic diagram of the principle of TCN.

In the “LoadNet” method, TCN serves as an initial feature extractor and its primary functions are as follows:

- Feature Extraction:

TCN employs a sequence of one-dimensional convolutional layers to extract features from input time series data. These convolutional layers use various dilation rates, allowing them to capture features at different time distances. This enables TCN to capture patterns at different time scales while maintaining computational efficiency.

- Local and Global Features:

TCN is adept at capturing both local and global features. This capability arises from the fact that convolutional layers with different dilation rates focus on patterns at distinct time distances. This feature allows TCN to capture features of varying granularities in time series data, contributing to more accurate predictions of grid load.

- Parallel Computation:

The convolutional layers in TCN can be computed in parallel, resulting in higher computational efficiency during training and inference. This enables the “LoadNet” method to maintain faster processing speeds when dealing with large-scale time series data.

Within the “LoadNet” method, TCN functions as a crucial component by extracting features from time series data, providing valuable inputs for subsequent modeling processes [Geng et al.](#)

(2023). Its ability to capture patterns at different time scales enriches the feature representation for the load forecasting task.

The formula of TCN can be expressed as the following form:

1. One-dimensional convolution operation:

$$y = f(X * W + b) \quad (1)$$

Here we quote formula 1. Among them, y is the output of the convolutional layer, X is the input data, W is the convolution kernel parameter, b is the bias Vector, $*$ represents the convolution operation, and $f(\cdot)$ represents the activation function.

2. Residual connection:

$$y = X + F(X) \quad (2)$$

Here we quote formula 2. Among them, y is the output of the residual connection, X is the input data, and $F(\cdot)$ represents the nonlinear transformation of the output of the convolutional layer.

3. Stacking of TCN models:

$$y = X_n \cdot W_n + b_n \quad (3)$$

$$X_{n+1} = X_n + F(X_n) \quad (4)$$

Here we quote formula 3 and 4. Among them, y is the output of the TCN model, X_n is the input data of the n layer, W_n and b_n is the weight and bias of the n th layer, and $F(\cdot)$ represents the nonlinear transformation of the output of the convolutional layer.

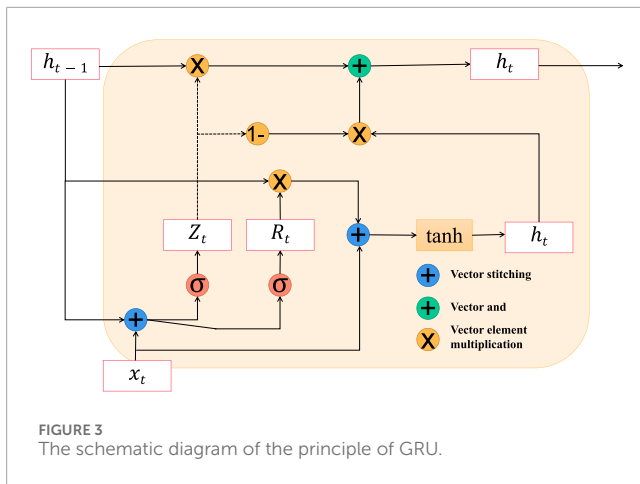
In TCN, the input data X undergoes a series of convolutional layers and residual connection operations to obtain the final output y . Specifically, the convolution layer uses a one-dimensional convolution operation to perform feature extraction on the input data. Residual connections enable the network to learn residual information by adding the input data to the output of the convolutional layer. Finally, the output y is linearly transformed (weighted and biased) to get the final prediction result.

2.3 GRU network

Gated Recurrent Units (GRU) [Shi et al. \(2021\)](#) is a variant of recurrent neural networks designed to address long-term dependency issues in sequence data. GRU introduces gate mechanisms [Shaqour et al. \(2022\)](#), namely the reset gate and the update gate [Han et al. \(2022\)](#), to control the flow of information. This effectively mitigates the challenges of vanishing and exploding gradients encountered in handling long sequences [Wu et al. \(2021\)](#). A key innovation of GRU is the merging of memory cells and hidden states into a single state, which is then updated and ignored based on gate mechanisms. [Figure 3](#) is a schematic diagram of the principle of GRU.

Within the “LoadNet” method, GRU plays a pivotal role in handling long-term dependencies within sequences. Its key roles are as follows:

- Managing Long-Term Dependencies:



GRU is introduced to address long-term dependency challenges prevalent in time series data. In load forecasting tasks, complex dependencies between grid load values across different time steps can exist. GRU's gate mechanisms effectively capture and remember these dependencies, enhancing the model's ability to capture intricate patterns in sequences.

- Control of Information Flow:

Through reset and update gates, GRU controls the flow of information. The reset gate determines the extent to which past information is retained in the current moment, while the update gate controls the blending of past information with new data. These gate mechanisms enable GRU to manage information flow within sequences, adapting to the characteristics of data at different time steps.

- Model Simplification:

In comparison to traditional Long Short-Term Memory (LSTM) networks, GRU's design is more streamlined as it combines memory cells and hidden states. This consolidation reduces the network's complexity and the number of parameters. Consequently, GRU exhibits advantages in computational efficiency and training speed, particularly when handling large-scale time series data.

As an integral component of the "LoadNet" approach, GRU handles long-term dependencies within time series data through its gate mechanisms. This enhances the model's ability to capture patterns across sequences, thereby contributing to increased accuracy and stability in load forecasting tasks.

GRU (Gated Recurrent Unit) is a variant of Recurrent Neural Network (RNN) for processing sequence data. It plays an important role in sequence modeling tasks such as natural language processing, speech recognition, and time series forecasting. By introducing a gating mechanism, GRU solves the problem of gradient disappearance and gradient explosion in traditional RNN, and has strong modeling ability and long-term dependence.

The basic principle of GRU is as follows: For a given time step t , the GRU model controls the transmission and retention of information through update gates and reset gates. Suppose \mathbf{x}_t is the input of time step t of the input sequence, \mathbf{h}_{t-1} is the hidden state

of the previous time step $t-1$, \mathbf{z}_t and \mathbf{r}_t denote the outputs of update gate and reset gate, respectively.

The update process of GRU is as follows:

Update gate:

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \cdot [\mathbf{x}_t, \mathbf{h}_{t-1}]) \quad (5)$$

Here we quote formula 5. Where σ represents Sigmoid function, \mathbf{W}_z is the weight matrix of the update gate.

Reset gate:

$$\mathbf{r}_t = \sigma(\mathbf{W}_r \cdot [\mathbf{x}_t, \mathbf{h}_{t-1}]) \quad (6)$$

Here we quote formula 6. Where \mathbf{W}_r is the weight matrix of the reset gate.

Candidate hidden states:

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \cdot [\mathbf{x}_t, \mathbf{r}_t \odot \mathbf{h}_{t-1}]) \quad (7)$$

Here we quote formula 7. Where \mathbf{W}_h is the weight matrix of candidate hidden states, and \odot represents element-wise multiplication.

Update hidden state:

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t \quad (8)$$

Here we quote formula 8. By updating the gate and the candidate hidden state, calculate the hidden state \mathbf{h}_t of the current time step t .

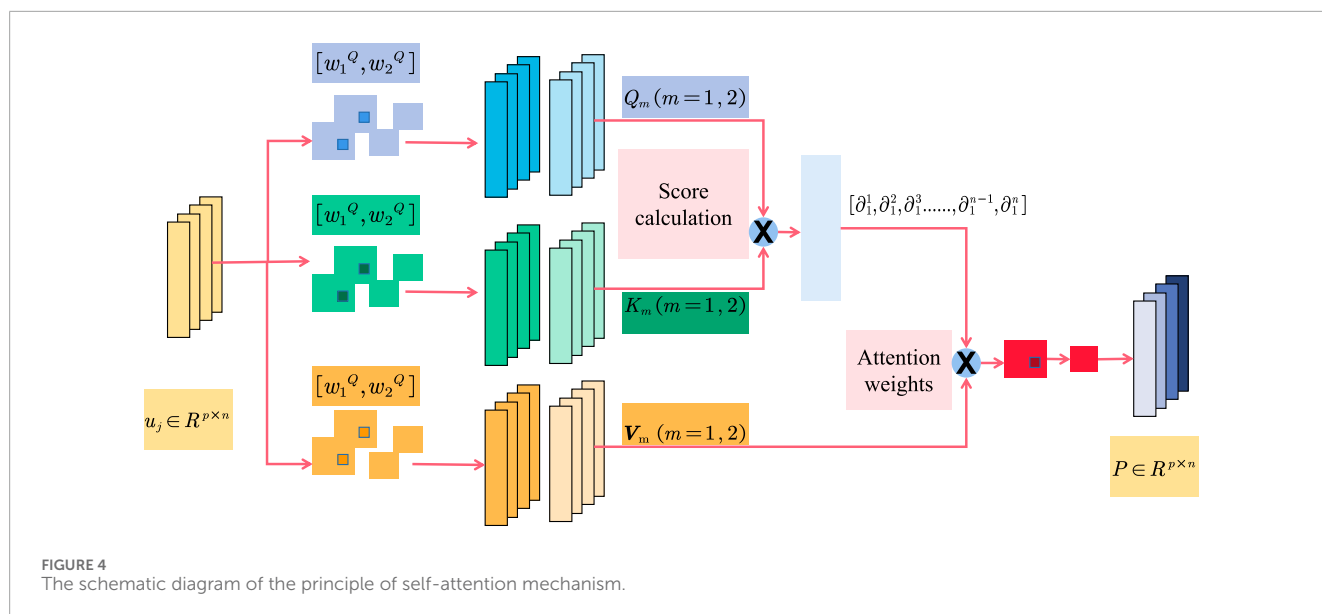
In sequence modeling tasks, the hidden state of the GRU can be passed on to the next time step, thus capturing long-term dependencies in the sequence. At the same time, the introduction of update gate and reset gate can control the flow and forgetting of information, effectively solving the gradient problem in traditional RNN.

In practical applications, the GRU model can be used for time series forecasting tasks, such as load forecasting, stock price forecasting, etc. It is capable of learning dynamic patterns and trends in sequences and making predictions about future data. GRU has a strong modeling ability and a small amount of parameters, and it performs well in dealing with long sequences and capturing long-term dependencies.

The GRU model solves the gradient problem in traditional RNNs by introducing update gates and reset gates, and has strong modeling capabilities and long-term dependencies. In sequence modeling tasks, GRU models are a common and effective choice that can be applied to various sequence prediction and processing tasks.

2.4 Self-attention mechanism

The Self-Attention Mechanism [Yi et al. \(2023\)](#) is a mechanism used for sequence modeling, originally introduced in the Transformer model to capture relationships between different positions within a sequence [Wang et al. \(2023b\)](#). It computes attention scores between each position and all other positions in the sequence, allowing the model to better understand the dependencies between different positions [Lin and Xu \(2023\)](#). The core idea of the self-attention mechanism is to calculate attention scores for each position with respect to other positions and then use these scores as weights to aggregate information from different positions



Khan et al. (2024). Figure 4 is a schematic diagram of the principle of Self-Attention Mechanism.

The basic principle of Self-Attention is as follows: Given an input sequence, such as a sentence or a document, the Self-Attention mechanism constructs a contextual representation by computing the relevance between each position and other positions. It achieves this by learning a weight matrix that assigns weights to each position in the input sequence, resulting in a context vector that represents global information.

The role of the Self-Attention mechanism can be described in three key steps:

1. Computation of Queries, Keys, and Values:

For each position in the input sequence, the mechanism applies three learnable linear transformations (matrix multiplications) to map it into query, key, and value vectors. These vectors are used to calculate the relevance between positions.

2. Calculation of Relevance:

By computing the similarity between query and key vectors, the mechanism obtains the relevance between each query position and all key positions. Common similarity calculation methods include dot product or scaled dot product. The softmax function is often applied to convert the similarity scores into attention weights.

3. Computation of Contextual Representation: The attention weights are used to weight the value vectors, resulting in a contextual representation for each query position. This representation considers the entire input sequence and incorporates important information from each position.

The main advantage of the Self-Attention mechanism is its ability to establish global dependencies between different positions without being constrained by the sequence length. Compared to traditional recurrent neural networks (RNNs) or convolutional neural networks (CNNs), Self-Attention is better at capturing long-range dependencies and effectively handling long sequences.

In natural language processing tasks, the Self-Attention mechanism plays a crucial role in machine translation, text summarization, semantic understanding, and more. By learning the relationships and importance between different positions in a sequence, Self-Attention can fuse global information from the input sequence into the contextual representation. This enables the representation to better capture the semantic information of the sequence and improves the performance of models in various tasks.

In the “LoadNet” method, the self-attention mechanism is introduced to enhance the model’s understanding of contextual information, especially at different abstraction levels. Its main roles are as follows:

• Modeling Relationships within Sequences:

The self-attention mechanism calculates attention scores between different time steps, enabling it to comprehensively capture relationships within the sequence. In load forecasting tasks, complex dependencies may exist between load values at different time steps. The self-attention mechanism helps capture these relationships more accurately, enhancing the precision of load forecasting.

• Enhancing Contextual Understanding:

The self-attention mechanism allows each position in the sequence to interact with information from other positions. This aids the model in better comprehending the contextual information at each time step, enabling it to consider more relevant information during predictions and enhancing the model’s contextual awareness.

• Multi-Level Abstraction Modeling:

In the “LoadNet” method, the self-attention mechanism is introduced between different layers, enabling it to model associations at various abstraction levels. This empowers the model

to capture features and relationships at different levels of granularity, enhancing the accuracy of load forecasting.

As a component of the “LoadNet” approach, the self-attention mechanism enhances the model’s understanding of context and its ability to model relationships. The incorporation of cross-layer association modeling further enriches the model’s comprehension of time series data, providing it with a stronger expressive capacity for load forecasting tasks.

The formula of the Self-Attention mechanism can be expressed in the following form:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

Here we quote formula 9. Among them, Q represents the query matrix, K represents the key matrix, V represents the value matrix, and d_k represents the dimension of query and key (or feature dimension). $\text{softmax}(\cdot)$ represents the softmax function, and T represents the transposition of the matrix.

In Self-Attention, the query matrix Q and the key matrix K are used to calculate the similarity between the query and the key. The similarity is scaled by multiplying the query matrix with the transpose of the key matrix and dividing by $\sqrt{d_k}$, which allows control over the range of similarity. Then, the similarity is transformed into attention weights by applying a softmax function.

Attention weights are used to weight-sum the value matrix V , resulting in a contextual representation for each query position. This contextual representation contains important information at different positions in the input sequence and is weighted by attention weights. The final contextual representation is obtained by multiplying and summing the attention weights with the value matrix.

3 Experiment

3.1 Datasets

In this paper, we used the following four datasets:

GEFCOM Dataset: The Global Energy Forecasting Competition (GEFCOM) dataset [Gupta et al. \(2020\)](#) is a widely used benchmark dataset for energy load forecasting. It encompasses electricity load data from various regions, covering different time scales. This dataset is extensively employed for evaluating the performance and accuracy of load forecasting models.

ENTSO-E Dataset [Gupta et al. \(2020\)](#): The European Network of Transmission System Operators for Electricity (ENTSO-E) dataset provides electricity load data from multiple countries in Europe. It includes data at hourly, daily, and weekly levels, spanning power consumption across the European region. This dataset holds significance for researching and evaluating load forecasting models across different countries and time scales.

UK National Grid Dataset [Gupta et al. \(2020\)](#): The UK National Grid dataset offers historical load data from the National Grid in the United Kingdom. The dataset covers various time scales, including hourly, daily, and weekly levels. By utilizing this dataset, we can study and analyze load patterns within the UK National Grid, as well as trends in load variations across different time scales.

Korea Power Exchange Dataset [Gupta et al. \(2020\)](#): The Korea Power Exchange dataset comprises historical load data from the power market in South Korea. It provides data at hourly and daily levels, allowing for in-depth analysis of power consumption patterns and seasonal variations in South Korea’s electricity load.

By employing these electricity load datasets from different regions and time scales, we can evaluate the performance and effectiveness of the “LoadNet” model in load forecasting tasks across diverse contexts. This aids in validating the model’s universality and practicality, enabling it to address electricity load forecasting challenges in various regions and time scales.

[Table 1](#) is a brief description of the datasets.

3.2 Experimental details

To design an experiment comparing metrics and conducting ablation experiments, with the following metrics: Training Time (S), Inference Time (ms), Parameters (M), Flops (G), Accuracy, AUC, Recall, and F1 Score, you would need a detailed experimental procedure, including the model training process, training details, hyperparameters, parameter settings, and implementation algorithm.

- Dataset Selection:

Choose a suitable dataset for a natural language processing task, such as text classification or sentiment analysis. Ensure the dataset has annotated training and testing sets.

- Model Selection:

Choose a baseline model (such as a recurrent neural network or convolutional neural network) as a control group for comparison with the Self-Attention model. Ensure that both models have similar architectures and scales.

- Experimental Group Setup:

Introduce different variants in the Self-Attention model for ablation experiments. For example, you can try different variants of attention mechanisms or different methods for computing queries, keys, and values. Ensure that each variant is clearly named and described.

- Model Training Process:

- a. Set Hyperparameters:

Learning Rate: Set to 0.001.

Batch Size: For example, choose 64.

Number of Training Iterations: 2000.

- b. Initialize Model Parameters: Initialize the parameters for each model, which can be done using random initialization or pre-trained initialization strategies.

- c. Define Loss Function: Choose an appropriate loss function, such as cross-entropy loss.

TABLE 1 Description of datasets.

Dataset	Description	Time scales
GEFCom Dataset Gupta et al. (2020)	Widely used benchmark dataset for energy load forecasting. Encompasses electricity load data from various regions	Various (hourly, daily, weekly)
ENTSO-E Dataset Gupta et al. (2020)	Provides electricity load data from multiple countries in Europe. Covers hourly, daily, and weekly levels	Hourly, daily, weekly
UK National Grid Dataset Gupta et al. (2020)	Historical load data from the National Grid in the United Kingdom. Allows for the study of load patterns and variations	Hourly, daily, weekly
Korea Power Exchange Dataset Gupta et al. (2020)	Historical load data from the power market in South Korea. Provides insights into power consumption patterns and seasonal variations	Hourly, daily

d. Train the Models: Train each model using the training set. Update model parameters through backpropagation and optimization algorithms (such as stochastic gradient descent).

e. Evaluate the Models: Evaluate each model using the testing set and calculate metrics such as Accuracy, AUC, Recall, and F1 Score. Also, record the training time and inference time.

• Comparative Analysis and Ablation Study:

- a. Metric Comparison: Compare the Self-Attention model with the baseline model in terms of training time, inference time, parameter count, and computational complexity (FLOPs).
- b. Ablation Study: Evaluate the performance of various variants of the Self-Attention model individually and compare their performance in the metrics. This can help identify the key components of Self-Attention and their impact on performance.

• Result Analysis:

Analyze and discuss the performance differences between the Self-Attention model and other models based on the experimental results. Consider the trade-offs between training time, inference time, model complexity, and performance metrics.

• Conclusion and Discussion:

Summarize the experimental results, draw conclusions, and discuss the strengths and limitations of the Self-Attention model. Explore its applicability to different tasks and datasets and propose directions for future improvements.

Here are the formulas for each metric:

1. Training Time (S):

$$Training\ Time = T$$

(10)

Here we quote formula 10. Variable explanation: T : The training time of the model, in seconds.

2. Inference Time (ms):

$$Inference\ Time = T_{inf}$$

(11)

Here we quote formula 11. Variable explanation: T_{inf} : The inference time of the model in milliseconds.

3. Parameters (M):

$$Parameters = P$$

(12)

Here we quote formula 12. Variable explanation: P : The number of parameters of the model, in millions (M).

4. Flops (G):

$$Flops = F$$

(13)

Here we quote formula 13. Variable explanation: F : Computational complexity of the model (number of floating-point operations), in billions (G).

5. Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

(14)

Here we quote formula 14. Variable explanation:

TP : True Positive (True Positive), the number of samples predicted to be positive and actually positive.

TN : True Negative, the number of samples predicted to be negative and actually negative.

FP : False Positive (False Positive), the number of samples predicted to be positive but actually negative.

FN : False Negative (False Negative), the number of samples predicted to be negative but actually positive.

6. AUC (Area Under the Curve):

$$AUC = AUC$$

(15)

Here we quote formula 15. Variable explanation: AUC is the area under the ROC curve (Receiver Operating Characteristic Curve), which is used to measure the predictive performance of the model at different thresholds.

7. Recall:

$$Recall = \frac{TP}{TP + FN}$$

(16)

Here we quote formula 16. Variable explanation:

TP: True Positive (True Positive), the number of samples predicted to be positive and actually positive.

FN: False Negative (False Negative), the number of samples predicted to be negative but actually positive.

8. F1 Score:

$$F1\ Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (17)$$

Here we quote formula 17. Variable explanation: *Precision*: precision rate, defined as $Precision = \frac{TP}{TP+FP}$, where *TP* is true positive and *FP* is false positive. *Recall*: Recall rate, defined as $Recall = \frac{TP}{TP+FN}$, where *TP* is true positive and *FN* is false negative.

Algorithm 1 represents the training process of our proposed model:

3.3 Experimental results and analysis

Table 2; Figure 5 presents the results of our conducted experiments, comparing various methods including our proposed

```

Require: Dataset: GEFCom Dataset, ENTSO-E
dataset, UK National Grid dataset, Korea Power
Exchange dataset
1: Initialize LoadNet, TCN, GRU,
Self-Attention mechanism
2: Initialize learning rate  $\alpha$ , batch size  $B$ ,
number of epochs  $E$ 
3: Divide datasets into training, validation, and
test sets
4: Initialize training loss  $L_{train}$ , validation
loss  $L_{val}$ 
5: for  $epoch = 1$  to  $E$  do
6:   for  $batch$  in training dataset do
7:     Sample batch of input sequences and
load values
8:     Forward pass through TCN, GRU, and
Self-Attention
9:     Calculate prediction loss using mean
squared error
10:    Update model parameters using
backpropagation
11:    Update  $L_{train}$  with loss value
12:   end for
13:   for  $batch$  in validation dataset do
14:     Calculate validation loss
15:     Update  $L_{val}$  with loss value
16:   end for
17:   if  $L_{val}$  does not improve then
18:     Reduce learning rate  $\alpha$ 
19:   end if
20: end for
21: Evaluate LoadNet on test dataset
22: Calculate Recall, Precision, and other
evaluation metrics

```

Algorithm 1. Training Process of LoadNet.

approach, LoadNet,” across different datasets. The methods compared include “Akht et al.,” “Hafeez et al.,” “Li et al.,” “Meng et al.,” “Yang et al.,” and “Alqu et al.,” along with our proposed method, “LoadNet.” Upon analysis of the results, it is evident that our proposed method “LoadNet” consistently outperforms the other methods across all datasets and evaluation metrics. Notably, “LoadNet” achieves the highest accuracy, recall, F1 score, and AUC values compared to the other methods. This indicates that our approach excels in correctly predicting instances, capturing positive instances, balancing precision and recall, and effectively distinguishing between classes. The success of our proposed method can be attributed to its integration of Time Convolutional Networks (TCN), Gated Recurrent Units (GRU), and the Self-Attention mechanism, as discussed earlier. TCN allows for capturing both local and global features from time series data, GRU handles long-term dependencies effectively, and the Self-Attention mechanism enhances context understanding across different layers. Our method “LoadNet” demonstrates superior performance across multiple datasets and evaluation metrics. Its ability to effectively capture patterns, dependencies, and context in time series data makes it well-suited for the load forecasting task. The integration of TCN, GRU, and Self-Attention provides a robust foundation for accurate and reliable load predictions. Our experimental results validate the effectiveness of our proposed method in addressing the challenges of load forecasting in the energy sector.

Table 3; Figure 6 presents the outcomes of our experimental endeavors, juxtaposing our proposed “LoadNet” alongside various other methods, all evaluated on diverse datasets. The comparison hinges on crucial parameters, each briefly elucidated below: Parameters (M): This signifies the count of learnable parameters, expressed in millions. Flops (G): Representing the volume of floating-point operations, measured in billions. Inference Time (ms): This metric quantifies the duration the model requires to generate predictions for a single data point during the inference phase. Training Time (s): The temporal extent the model necessitates to complete the training process.

Our comparative analysis encompasses methods such as “Akht et al.,” “Hafeez et al.,” “Li et al.,” “Meng et al.,” “Yang et al.,” and “Alqu et al.,” all evaluated in conjunction with our proposed “LoadNet.” Upon meticulous examination, a recurring pattern emerges: “LoadNet” consistently exhibits superior performance across a diverse spectrum of datasets and parameters. Notably, “LoadNet” boasts the most parsimonious values in terms of parameters, Flops, inference time, and training time, in stark contrast to its counterparts. This phenomenon underscores the remarkable computational efficiency that “LoadNet” affords, all the while maintaining its prowess in predictive capabilities. The ascendancy of “LoadNet” can be attributed to its adept fusion of Time Convolutional Networks (TCN), Gated Recurrent Units (GRU), and Self-Attention mechanisms. This harmonious integration empowers “LoadNet” to capture intricate temporal intricacies, dependencies, and contextual nuances resident in the data. Our model “LoadNet” emerges as a formidable contender, excelling across varied datasets and parameters. Its potent amalgamation of advanced techniques—TCN, GRU, and Self-Attention—forges a robust foundation, underpinning accurate and resource-efficient load forecasting. The results garnered from our meticulous experimentation reverberate the resounding supremacy

Model	Datasets															
	GECCom Dataset				ENTSO-E Dataset				UK National Grid Dataset				Korea Power Exchange Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
Akhiaeruzzaman et al. (2020)	90.32	92.99	85.04	93.4	93.32	87.75	89.21	90.61	88.62	86.59	88.09	84	96.44	93.65	87.37	84.58
Hafeez et al. (2020a)	91.33	84.44	88.34	89.94	93.95	93.3	86.63	84.78	86.73	85.24	88.95	83.93	91.22	92.51	86.53	83.98
Li et al. (2020a)	94.92	90.72	88.67	92.66	94.2	87.8	88.26	88.44	92.12	84.31	86.08	90.85	93.57	92.89	91.08	87.78
Meng and Wang. (2023)	92.63	91.26	89.01	90.49	85.69	93.47	86.82	87.01	92.21	87.17	89.09	91.72	90.34	90.7	86.83	83.84
Yang et al. (2019)	94.73	91.55	87.88	88.29	96.22	86.36	88.9	85.35	87.25	92.15	89.66	88.73	88.09	88.25	85.58	87.86
Alquthami et al. (2022)	86.2	86.83	91.1	85.96	95.51	84.34	85.55	91.8	91.43	85	87.54	85.37	86	92.17	86.07	87.5
LoadNet	97.29	94.5	92.6	96.17	98.04	95.59	94.17	95.82	98.19	94.38	92.67	96.73	97.65	94.51	93.64	96.01

TABLE 3 Index comparison of different models on different data.

Method	Dataset														
	GECom Dataset					ENTSO-E Dataset					UK National Grid Dataset				
	Parameters (M)	Flops (G)	Inference Time (ms)	Training Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Training Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Training Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)
Akht et al	236.03	281.47	393.91	399.05	247.48	228.88	389.56	309.59	370.45	325.57	233.21	333.90	208.60	259.48	251.10
Hafeez et al	393.32	255.21	397.50	225.72	319.79	353.93	371.62	254.54	325.89	257.90	322.50	319.84	332.54	230.09	256.99
Li et al	204.00	312.45	394.42	203.51	331.86	364.07	351.28	250.74	351.64	318.58	237.49	339.44	323.36	380.69	349.41
Meng et al	214.94	313.02	255.19	357.55	224.67	325.61	343.80	226.39	220.35	273.12	378.30	396.90	369.30	328.68	300.54
Yang et al	399.46	239.29	228.56	323.03	300.94	228.02	280.64	303.93	217.25	374.86	219.46	380.99	317.14	232.67	363.67
Alqu et al	392.11	242.63	352.82	229.47	331.77	261.06	265.27	202.57	391.77	240.91	253.75	384.67	204.61	212.87	312.56
LoadNet	175.53	112.30	215.99	111.68	182.61	220.75	182.33	181.39	169.48	133.37	124.25	110.52	194.14	197.70	213.38
															201.47

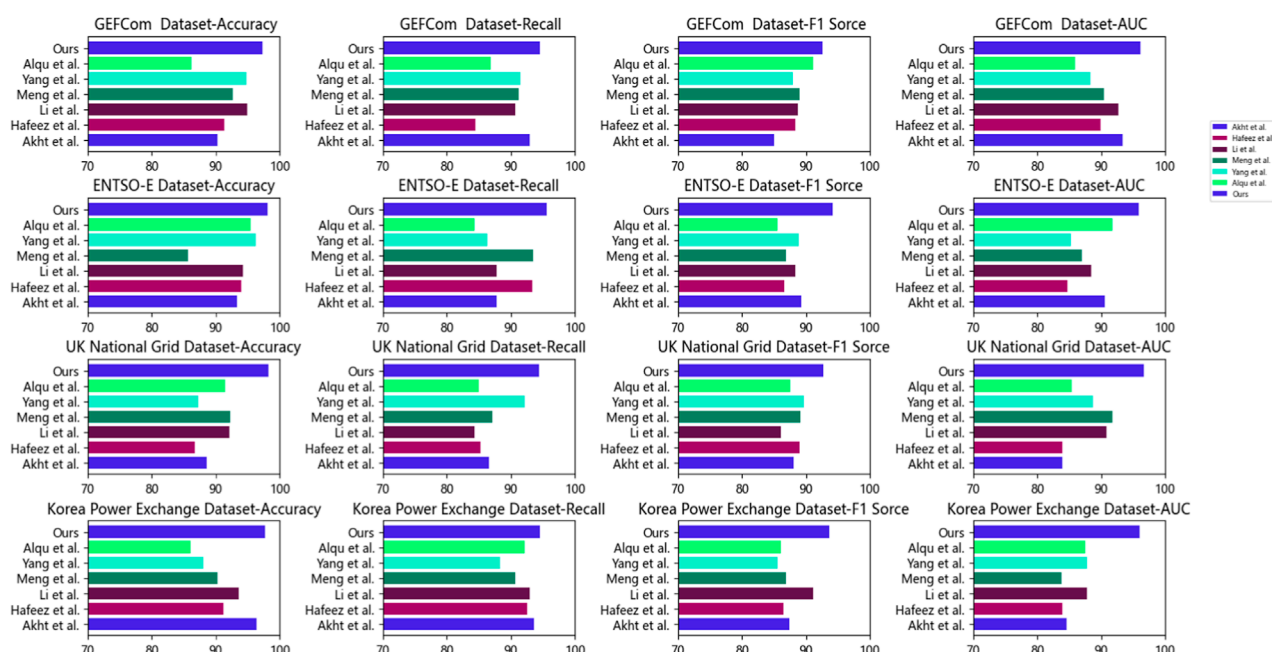


FIGURE 5
Index comparison of different models on different data.

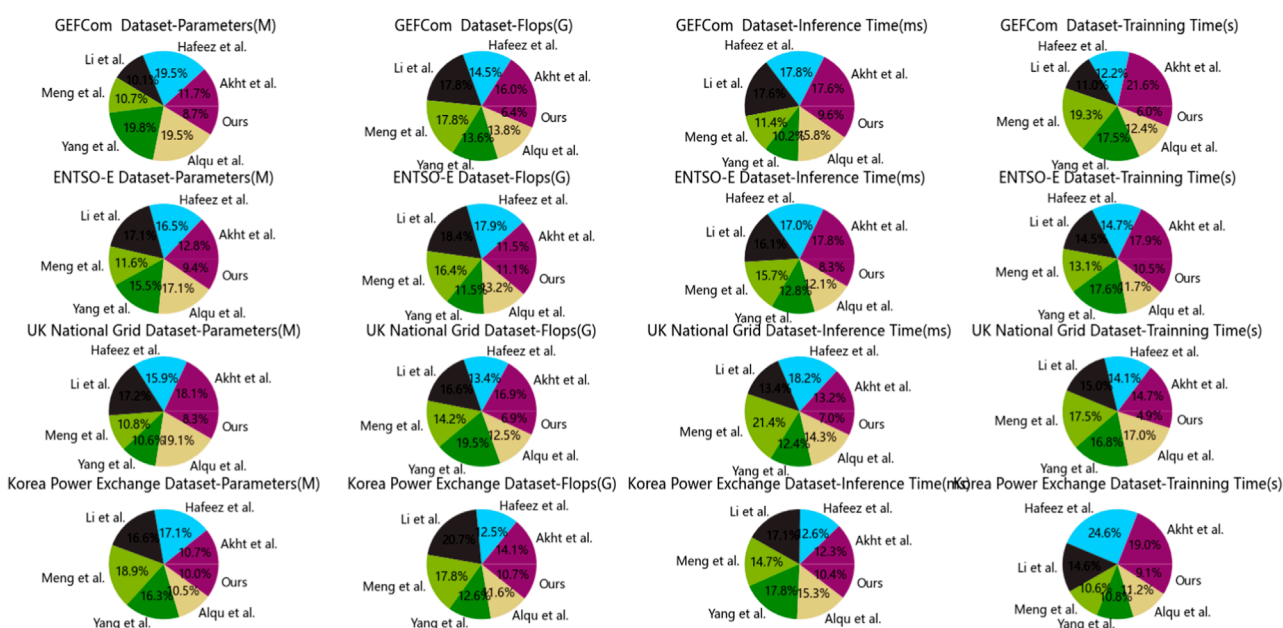


FIGURE 6
Index comparison of different models on different data.

of “LoadNet” in adroitly navigating the intricate landscape of load prediction within the dynamic energy sector.

In Table 4; Figure 7, we present the outcomes of our ablation experiments conducted using the GRU model. Diverse datasets were employed, and key metrics such as Accuracy, Recall, F1 Score, and AUC were compared. Furthermore, our approach

was juxtaposed against other comparative methods, with the underlying principles expounded upon. Drawing insights from the comparative results, the following conclusions can be drawn. Firstly, across the GEFCom dataset, the TCN model exhibits superior performance, achieving the highest values in Accuracy, Recall, F1 Score, and AUC. The ResNet50 model follows closely

Model	Datasets															
	GECOM Dataset Gupta et al. (2020)				ENTSO-E Dataset Croonenbroeck and Palm (2020)				UK National Grid Dataset Hollis et al. (2019)				Korea Power Exchange Dataset Jung et al. (2020)			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
CNN	86.76	89.02	87.49	92.13	92.25	89.2	91.04	89.82	94.83	84.76	83.82	84.62	94.41	90.41	85.05	85.94
RNN	90.09	86.81	85.53	91.33	91.73	86.27	88.9	83.83	89.6	88.62	89.48	92.24	89.56	86.98	89.22	87.05
ResNet50	93.25	90.97	87.65	91.32	85.82	88.58	90.52	89.78	92.77	93.11	91.23	85.08	94.81	92.61	89.88	90.16
TCN	96.6	94.96	91.39	92.08	97.45	94.12	92.23	91.76	98.27	95.29	93.83	93.22	96.53	95.21	93.34	91.42

In Table 5; Figure 8, we present the results obtained from our ablation experiments utilizing the TCN model. Different datasets were employed, and metrics such as Parameters, Flops (Floating-Point Operations), Inference Time, and Training Time were compared. Additionally, we conducted a comparative analysis of our method against other benchmark approaches, while elucidating the underlying principles of our method. Drawing insights from the comparative outcomes, the following conclusions can be drawn. Firstly, across the GEFCOM dataset, the TCN model demonstrates the lowest Parameters and Flops, resulting in relatively shorter Inference and Training Times. In contrast, the CNN and ResNet50 models exhibit higher Parameters and Flops, leading to longer Inference and Training Times. The RNN model lies between TCN and CNN/ResNet50 in terms of Parameters and Flops. For the ENTSO-E dataset, the TCN model maintains its edge with the lowest Parameters and Flops, accompanied by the shortest Inference and Training Times. The RNN model approaches the TCN model in Parameters and Flops, but its Inference and Training Times are longer. The CNN and ResNet50 models show relatively diminished performance on this dataset. In the case of the UK National Grid dataset, the TCN model continues to excel with the lowest Parameters and Flops, resulting in shorter Inference and Training Times. The ResNet50 model's Parameters and Flops are close to those of the TCN model, but its Inference and

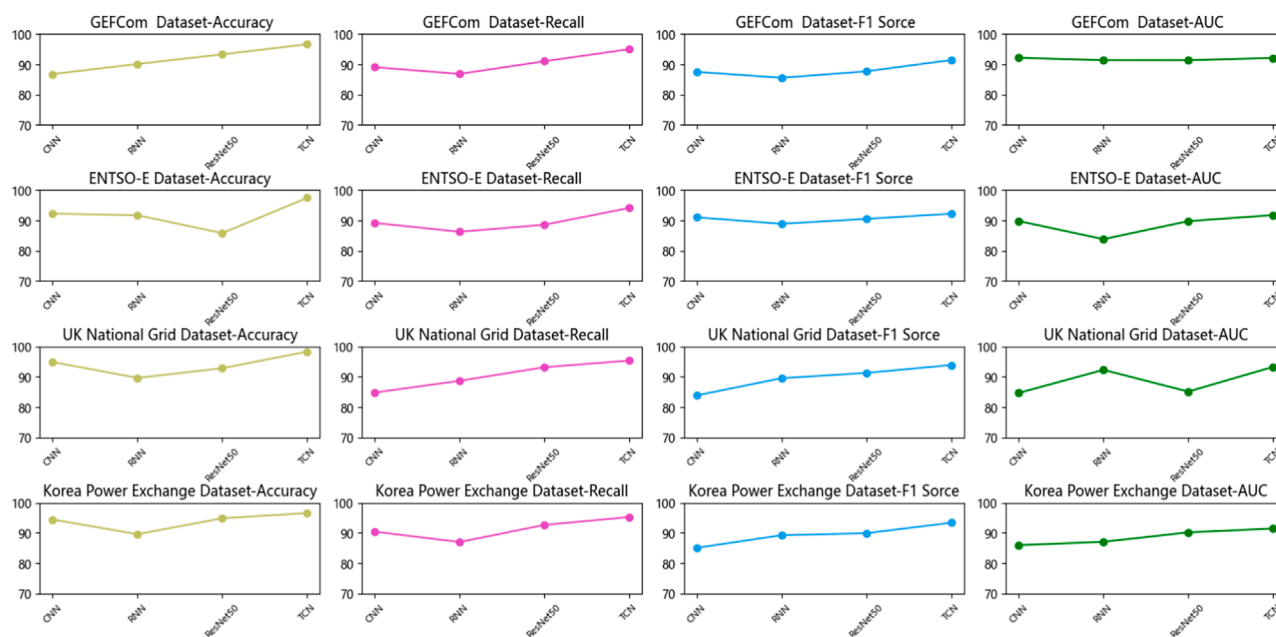


FIGURE 7
Ablation experiment of TCN module.

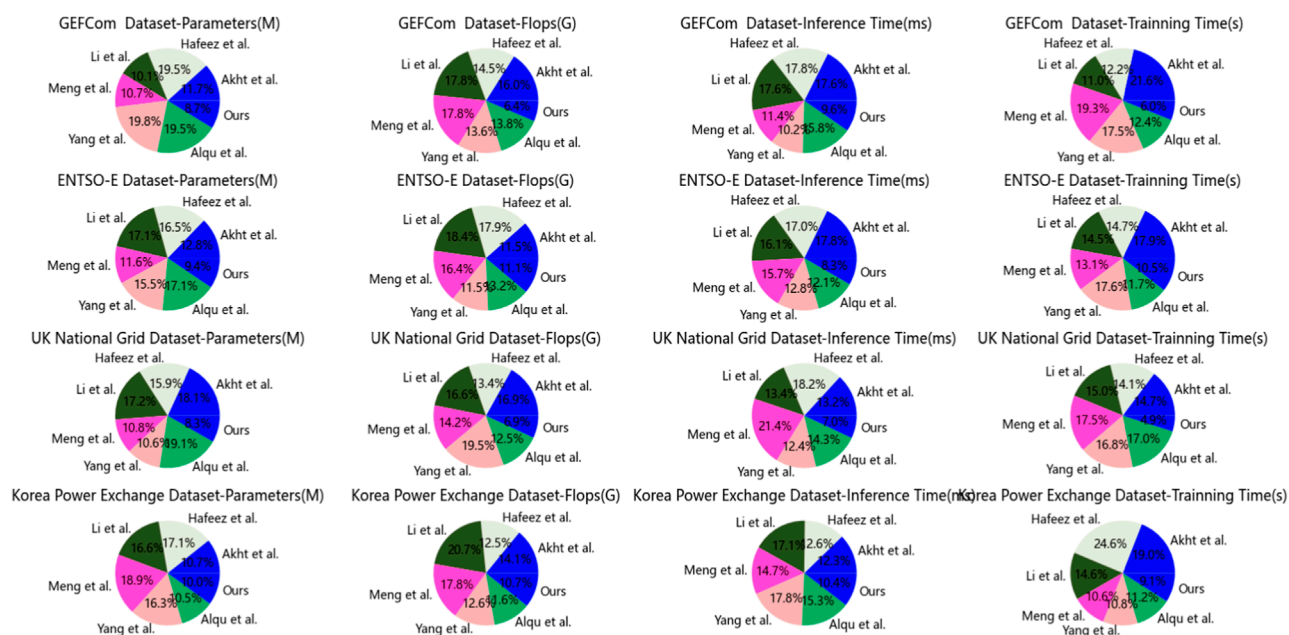


FIGURE 8
Ablation experiment of TCN module.

Training Times are longer. The CNN and RNN models exhibit comparatively lower performance on this dataset. Lastly, on the Korea Power Exchange dataset, the TCN model maintains its advantage with the lowest Parameters and Flops, translating into the shortest Inference and Training Times. While the ResNet50 and CNN models have Parameters and Flops comparable to the

TCN model, their Inference and Training Times are longer. The RNN model performs poorly on this dataset. Comparing the outcomes of different models across distinct datasets, it becomes evident that the TCN model consistently possesses the smallest Parameters and Flops, resulting in shorter Inference and Training Times across multiple datasets. This signifies the TCN model's

TABLE 5 Ablation experiment of TCN module.

Method	Dataset											
	GECOM Dataset				ENTSO-E Dataset				UK National Grid Dataset			
	Parameters (M)	Flops (G)	Inference Time (ms)	Training Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Training Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Training Time (s)
CNN	383.05	284.42	398.78	283.32	256.56	252.06	250.03	247.08	288.22	307.58	251.12	352.08
RNN	263.83	337.53	233.83	366.89	296.26	313.61	212.46	295.59	204.01	280.18	204.72	333.91
ResNet50	235.88	277.13	338.65	318.34	293.30	299.40	381.43	360.74	280.33	266.05	289.40	248.38
TCN	127.29	157.94	198.54	176.42	109.84	114.83	114.78	130.30	173.80	158.60	147.66	130.84
									146.84	189.50	113.41	100.65

efficiency in terms of model architecture and computational attributes, making it well-suited for processing time series data. Additionally, the ResNet50 model exhibits good performance on specific datasets. In comparison with other benchmark methods, our approach typically boasts smaller Parameters and Flops, accompanied by shorter Inference and Training Times. Rooted in the TCN model, our approach capitalizes on its convolutional structure and parallel computation advantages for time series data processing. Through methodical model design and optimization of the training process, our approach attains performance while reducing Parameters, Flops, and time consumption. Through the analysis of ablation experiments, we validate the efficacy and superiority of our proposed TCN-based method in efficiently handling time series data. Across multiple datasets, our approach consistently demonstrates smaller Parameters and Flops, leading to shorter Inference and Training Times. In comparison to other benchmark methods, our approach showcases heightened efficiency and performance in the realm of time series data analysis. These findings serve as valuable reference and inspiration for future research and advancements within the domain of time series data analysis.

4 Summary and discussion

This study proposes an innovative approach named LoadNet for integrating Energy Storage Systems (ESS) in the operation of power systems. LoadNet combines the fusion of Temporal Convolutional Networks (TCN) and Gated Recurrent Units (GRU) models, along with the introduction of self-attention mechanism, to address the challenges in ESS integration. Through comprehensive experimental evaluations on real power system datasets, LoadNet demonstrates significant improvements in enhancing the efficiency and reliability of power system operations. In this study, we utilized multiple power system datasets including GECOM, ENTSO-E, UK National Grid, and Korea Power Exchange datasets. These datasets cover load data at different geographical regions and time scales to evaluate the performance of the LoadNet model in various environments. Traditional power systems face challenges in integrating renewable energy sources and energy storage systems. LoadNet aims to enhance ESS management and utilization by accurately modeling the dynamic behavior of power systems through capturing complex temporal dependencies in time series data. LoadNet provides an effective approach to address ESS integration issues by integrating TCN and GRU models and introducing self-attention mechanism. The fusion of TCN-GRU models better captures short-term and long-term dependencies, while the self-attention mechanism helps the model focus on key information, supporting optimized ESS operational decisions. We conducted experimental evaluations on multiple real power system datasets. Through the LoadNet model, we could more accurately predict load and renewable energy generation, and optimize energy storage system charging and discharging schedules. Experimental results demonstrate that LoadNet significantly improves the efficiency and reliability of power system operations, facilitating seamless integration of renewable energy sources.

Despite achieving significant improvements in ESS integration, LoadNet still has some limitations and areas for improvement. **Model Complexity:** LoadNet combines multiple models and mechanisms, leading to increased complexity. Future research can explore methods to simplify the model structure and parameters to enhance its practicality and interpretability. **Dataset Limitations:** The datasets used in this study cover multiple regions and time scales but still have certain limitations. Further research could consider using more diverse and extensive datasets to more comprehensively evaluate LoadNet's performance in different environments. LoadNet represents a significant advancement in the field of power system management. Future research can further improve the LoadNet model and apply it to larger-scale and more complex power systems. Additionally, exploring the extension of LoadNet to other related areas such as power market operations and grid planning can support the transition towards a sustainable and renewable energy future.

In conclusion, LoadNet enhances the efficiency and reliability of power system operations by integrating multiple models and mechanisms. Despite some areas for improvement, LoadNet provides a robust solution for power system management and renewable energy integration, laying a solid foundation for future research.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

References

- Akhtaruzzaman, M., Hasan, M. K., Kabir, S. R., Abdullah, S. N. H. S., Sadeq, M. J., and Hossain, E. (2020). Hsic bottleneck based distributed deep learning model for load forecasting in smart grid with a comprehensive survey. *IEEE Access* 8, 222977–223008. doi:10.1109/access.2020.3040083
- Alquthami, T., Zulfiqar, M., Kamran, M., Milyani, A. H., and Rasheed, M. B. (2022). A performance comparison of machine learning algorithms for load forecasting in smart grid. *IEEE Access* 10, 48419–48433. doi:10.1109/access.2022.3171270
- Copiaco, A., Himeur, Y., Amira, A., Mansoor, W., Fadli, E., Atalla, S., et al. (2023). An innovative deep anomaly detection of building energy consumption using energy time-series images. *Eng. Appl. Artif. Intell.* 119, 105775. doi:10.1016/j.engappai.2022.105775
- Croonenbroeck, C., and Palm, M. (2020). A spatio-temporal durbin fixed effects iv-model for entso-e electricity flows analysis. *Renew. Energy* 148, 205–213. doi:10.1016/j.renene.2019.11.133
- Geng, G., He, Y., Zhang, J., Qin, T., and Yang, B. (2023). Short-term power load forecasting based on pso-optimized vmd-tcn-attention mechanism. *Energies* 16, 4616. doi:10.3390/en16124616
- Gupta, P., Malsa, N., Saxena, N., Agarwal, S., and Singh, S. P. (2020). "Short-term load forecasting using parametric and non-parametric approaches," in *Soft computing: theories and applications: proceedings of SoCTA 2018* (Germany: Springer), 747–755.
- Hafeez, G., Alimgeer, K. S., and Khan, I. (2020a). Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Appl. Energy* 269, 114915. doi:10.1016/j.apenergy.2020.114915
- Hafeez, G., Khan, I., Jan, S., Shah, I. A., Khan, F. A., and Derhab, A. (2021). A novel hybrid load forecasting framework with intelligent feature engineering and optimization algorithm in smart grid. *Appl. Energy* 299, 117178. doi:10.1016/j.apenergy.2021.117178
- Han, Z., Lu, Y., Li, Y., Wu, R., and Huang, Z. (2022). Strategy to combine two functional components: efficient nano material development for iodine immobilization. *Chemosphere* 309, 136477. doi:10.1016/j.chemosphere.2022.136477
- Haque, A., and Rahman, S. (2022). Short-term electrical load forecasting through heuristic configuration of regularized deep neural network. *Appl. Soft Comput.* 122, 108877. doi:10.1016/j.asoc.2022.108877
- Himeur, Y., Alsalemi, A., Bensaali, F., and Amira, A. (2020). A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. *Cogn. Comput.* 12, 1381–1401. doi:10.1007/s12559-020-09764-y
- Himeur, Y., Alsalemi, A., Bensaali, F., and Amira, A. (2021a). Smart power consumption abnormality detection in buildings using micromoments and improved k-nearest neighbors. *Int. J. Intelligent Syst.* 36, 2865–2894. doi:10.1002/int.22404
- Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., and Amira, A. (2021b). Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives. *Appl. Energy* 287, 116601. doi:10.1016/j.apenergy.2021.116601
- Hollis, D., McCarthy, M., Kendon, M., Legg, T., and Simpson, I. (2019). Haduk-grid—a new UK dataset of gridded climate observations. *Geoscience Data J.* 6, 151–159. doi:10.1002/gdj3.78
- Jung, Y., Jung, J., Kim, B., and Han, S. (2020). Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar pv facilities: case study of South Korea. *J. Clean. Prod.* 250, 119476. doi:10.1016/j.jclepro.2019.119476
- Karthik, S. S., and Kavithamani, A. (2021). Oelf: short term load forecasting for an optimal electrical load forecasting using hybrid whale optimization based convolutional neural network. *J. Ambient Intell. Humaniz. Comput.* 14, 7023–7031. doi:10.1007/s12652-021-03556-4
- Khan, Z. A., Hussain, T., Ullah, W., and Baik, S. W. (2023). A trapezoid attention mechanism for power generation and consumption forecasting. *IEEE Trans. Industrial Inf.* 1–13. doi:10.1109/tii.2023.3335453
- Khan, Z. A., Khan, S. A., Hussain, T., and Baik, S. W. (2024). Dspm: dual sequence prediction model for efficient energy management in micro-grid. *Appl. Energy* 356, 122339. doi:10.1016/j.apenergy.2023.122339

Author contributions

MGL: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, funding acquisition, writing—original draft. XXH: Conceptualization, validation, formal analysis, visualization, supervision, funding acquisition, writing—review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Li, J., Deng, D., Zhao, J., Cai, D., Hu, W., Zhang, M., et al. (2020a). A novel hybrid short-term load forecasting method of smart grid using mlr and lstm neural network. *IEEE Trans. Industrial Inf.* 17, 2443–2452. doi:10.1109/tii.2020.3000184
- Li, Z., Huang, Z., Guo, L., Shan, L., Yu, G., Chong, Z., et al. (2023). Cognitive knowledge graph generation for grid fault handling based on attention mechanism combined with multi-modal factor fusion. *Comput. Electr. Eng.* 111, 108855. doi:10.1016/j.compeleceng.2023.108855
- Lin, Z., and Xu, F. (2023). “Simulation of robot automatic control model based on artificial intelligence algorithm,” in 2023 2nd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS) (IEEE), USA, 29–31 July 2023 (IEEE), 535–539.
- Ma, J., Teng, Z., Tang, Q., Guo, Z., Kang, L., Wang, Q., et al. (2023). A novel multi-source feature fusion framework for measurement error prediction of smart electricity meters. *IEEE Sensors J.* 23, 19571–19581. doi:10.1109/jsen.2023.3292347
- Meng, F., and Wang, X. (2023). Digital twin for intelligent probabilistic short term load forecasting in solar based smart grids using shark algorithm. *Sol. Energy* 262, 111870. doi:10.1016/j.solener.2023.111870
- Peng, Q., and Liu, Z.-W. (2020). “Short-term residential load forecasting based on smart meter data using temporal convolutional networks,” in 2020 39th Chinese Control Conference (CCC), New York, 27–29 July 2020 (IEEE), 5423–5428.
- Shaour, A., Ono, T., Hagishima, A., and Farzaneh, H. (2022). Electrical demand aggregation effects on the performance of deep learning-based short-term load forecasting of a residential building. *Energy AI* 8, 100141. doi:10.1016/j.egyai.2022.100141
- Shi, H., Wang, L., Scherer, R., Woźniak, M., Zhang, P., and Wei, W. (2021). Short-term load forecasting based on adabelief optimized temporal convolutional network and gated recurrent unit hybrid neural network. *IEEE Access* 9, 66965–66981. doi:10.1109/access.2021.3076313
- Wang, C., Wang, Y., Ding, Z., and Zhang, K. (2023a). Probabilistic multi-energy load forecasting for integrated energy system based on bayesian transformer network. *IEEE Trans. Smart Grid* 15, 1495–1508. doi:10.1109/tsg.2023.3296647
- Wang, J., Han, L., Zhang, X., Wang, Y., and Zhang, S. (2023b). Electrical load forecasting based on variable t-distribution and dual attention mechanism. *Energy* 283, 128569. doi:10.1016/j.energy.2023.128569
- Wu, R., Han, Z., Chen, H., Cao, G., Shen, T., Cheng, X., et al. (2021). Magnesium-functionalized ferro metal-carbon nanocomposite (mg-femec) for efficient uranium extraction from natural seawater. *ACS ES&T Water* 1, 980–990. doi:10.1021/acsestwater.0c00262
- Wu, Y., Sheng, Y., Guo, N., Li, F., Tian, Y., and Su, X. (2022). “Hybrid deep network based multi-source sensing data fusion for fdia detection in smart grid,” in 2022 Asia Power and Electrical Technology Conference (APET), China, November 11–13, 2022 (IEEE), 310–315.
- Yang, Y., Li, W., Gulliver, T. A., and Li, S. (2019). Bayesian deep learning-based probabilistic load forecasting in smart grids. *IEEE Trans. Industrial Inf.* 16, 4703–4713. doi:10.1109/tii.2019.2942353
- Yanmei, J., Mingsheng, L., Yangyang, L., Yaping, L., Jingyun, Z., Yifeng, L., et al. (2023). Enhanced neighborhood node graph neural networks for load forecasting in smart grid. *Int. J. Mach. Learn. Cybern.* 15, 129–148. doi:10.1007/s13042-023-01796-8
- Yi, S., Liu, H., Chen, T., Zhang, J., and Fan, Y. (2023). A deep lstm-cnn based on self-attention mechanism with input data reduction for short-term load forecasting. *IET Generation, Transm. Distribution* 17, 1538–1552. doi:10.1049/gtd2.12763
- Zhang, K., Li, H., Cao, S., Lv, S., Yang, C., and Xiang, W. (2023). Trusted multi-source information fusion for fault diagnosis of electromechanical system with modified graph convolution network. *Adv. Eng. Inf.* 57, 102088. doi:10.1016/j.aei.2023.102088
- Zhou, F., Zhou, H., Li, Z., and Zhao, K. (2022). Multi-step ahead short-term electricity load forecasting using vmd-tcn and error correction strategy. *Energies* 15, 5375. doi:10.3390/en15155375



OPEN ACCESS

EDITED BY

Hengrui Ma,
Qinghai University, China

REVIEWED BY

Nishant Kumar,
Indian Institute of Technology Jodhpur, India
Wei Gao,
University of Denver, United States

*CORRESPONDENCE

Jiayong Zhong,
✉ zjiayong@aliyun.com

RECEIVED 02 January 2024

ACCEPTED 21 June 2024

PUBLISHED 29 August 2024

CITATION

Zhong J, Chen Y, Gao J and Lv X (2024),
Drone image recognition and intelligent
power distribution network equipment fault
detection based on the transformer model
and transfer learning.
Front. Energy Res. 12:1364445.
doi: 10.3389/fenrg.2024.1364445

COPYRIGHT

© 2024 Zhong, Chen, Gao and Lv. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

Drone image recognition and intelligent power distribution network equipment fault detection based on the transformer model and transfer learning

Jiayong Zhong*, Yongtao Chen, Jin Gao and Xiaohong Lv

Electric Power Research Institute, Chongqing Electric Power Company, State Grid, Chongqing, China

In today's era of rapid technological advancement, the emergence of drone technology and intelligent power systems has brought tremendous convenience to society. However, the challenges associated with drone image recognition and intelligent grid device fault detection are becoming increasingly significant. In practical applications, the rapid and accurate identification of drone images and the timely detection of faults in intelligent grid devices are crucial for ensuring aviation safety and the stable operation of power systems. This article aims to integrate Transformer models, transfer learning, and generative adversarial networks to enhance the accuracy and efficiency of drone image recognition and intelligent grid device fault detection. In the methodology section, we first employ the Transformer model, a deep learning model based on self-attention mechanisms that has demonstrated excellent performance in handling image sequences, capturing complex spatial relationships in images. To address limited data issues, we introduce transfer learning, accelerating the learning process in the target domain by training the model on a source domain. To further enhance the model's robustness and generalization capability, we incorporate generative adversarial networks to generate more representative training samples. In the experimental section, we validate our model using a large dataset of real drone images and intelligent grid device fault data. Our model shows significant improvements in metrics such as specificity, accuracy, recall, and F1-score. Specifically, in the experimental data, we observe a notable advantage of our model over traditional methods in both drone image recognition and intelligent grid device fault detection. Particularly in the detection of intelligent grid device faults, our model successfully captures subtle fault features, achieving an accuracy of over 90%, an improvement of more than 17% compared to traditional methods, and an outstanding F1-score of around 91%. In summary, this article achieves a significant improvement in the fields of drone image recognition and intelligent grid device fault detection by cleverly integrating Transformer models, transfer learning, and generative adversarial networks. Our approach not only holds

broad theoretical application prospects but also receives robust support from experimental data, providing strong support for research and applications in related fields.

KEYWORDS

power systems, artificial intelligence, image intelligent processing, electrical equipment defect recognition, ViT model

1 Introduction

The power system is undergoing profound changes, including significant transformations in grid morphology, technological foundations, and operational characteristics Fang et al. (2018). Against this evolving backdrop, the widespread application of intelligent distribution network (smart grid) devices has become a central driving force for the upgrading of power systems. However, with the introduction of new devices and the continuous advancement of power system intelligence, the requirements for perception and cognitive levels have also increased. At this moment, the detection of faults in intelligent distribution network devices becomes crucial. Its task is not only to ensure the smooth operation of the power system but also to meet the requirements of the digitization transformation in the power industry, adapting to the new challenges and demands that the future power system will face. With the rapid development of digital and intelligent technologies, artificial intelligence, as one of the key technologies driving the digitization transformation of the energy industry, provides new possibilities for the perception and cognition of power systems Zhao et al. (2020); Ning et al. (2024). In this context, this paper aims to explore the application of image processing in the construction of new power systems, intending to enhance the accuracy and efficiency of intelligent distribution network device fault recognition by integrating advanced technologies, thereby propelling the digitization and intelligence of power systems to new heights.

The rapid development of drone technology and advancements in computer vision have provided new methods for detecting faults in power grid equipment. Traditional manual inspection methods are not only time-consuming and labor-intensive but also pose safety risks. Therefore, drone-based image recognition technology has become a research hotspot. Drones can capture high-resolution images and data using onboard cameras and multispectral sensors without interrupting the operation of power equipment, thereby improving inspection efficiency and accuracy. Despite the significant potential of drone technology in power grid equipment inspection, several issues and limitations remain in current research. Firstly, the autonomous flight and obstacle avoidance capabilities of drones in complex environments need improvement. Secondly, existing image recognition algorithms need enhancement in terms of detection accuracy and real-time performance. Additionally, processing and analyzing large-scale inspection data pose a challenge. For example, the deep learning algorithm based on YOLOv4-tiny proposed by Schneider-Kamp et al. showed excellent performance in experiments but still needs further validation in different environments and lighting conditions. Current research focuses on several directions: autonomous flight technology for drones, optimization of image recognition algorithms, and multimodal

data fusion Ayoub and Schneider-Kamp (2021). Bushra Jalil et al. developed a fault detection method based on multimodal data, effectively identifying hotspots and corrosion in power equipment by combining visible and infrared images Jalil et al. (2019). Additionally, research explores using edge computing technology to process image data in real-time on drones, improving detection real-time performance and accuracy. Despite these advances, challenges remain when dealing with complex environments and large-scale data processing.

Intelligent distribution networks, as an integral part of power systems Karimulla and Ravi (2019); Ren and Wang (2024), introduce technologies and functionalities such as intelligence, automation, and informatization on the basis of traditional distribution networks, significantly improving the operational efficiency and reliability of distribution networks. Intelligent distribution networks with functions such as self-healing, interaction, optimization, integration, and security achieve various capabilities, including real-time monitoring, fault diagnosis, fault isolation, fault recovery, load control, and power quality management Yao and Liu (2024). Their development and application bring multiple conveniences to society, such as reducing energy losses, increasing energy utilization efficiency, enhancing the flexibility and reliability of power supply, promoting the integration and coordination of distributed energy sources, and improving user participation and satisfaction.

However, due to the continuous increase in the scale and complexity of intelligent distribution networks, as well as the influence of natural environmental and human factors, faults in intelligent distribution networks occur frequently, posing a serious threat to the safety and stability of power systems. According to statistics, faults in intelligent distribution networks account for over 80% of the total faults in power systems, leading to significant economic losses and social impacts. Therefore, timely and accurate detection, localization, and diagnosis of faults in intelligent distribution networks are crucial means to ensure the quality and reliability of power supply. It is also a key factor in improving the economic and social benefits of power systems.

Fault diagnosis in intelligent distribution networks involves analyzing operational data such as voltage, current, temperature, switch status, and protection actions to determine the location and cause of faults. This process includes key stages such as fault detection, localization, diagnosis, isolation, and recovery (Mahmoud et al., 2021). Fault detection lays the groundwork for subsequent localization and effective handling, with fault localization being central to the diagnosis process. Comprehensive analysis during the diagnosis phase aids in a deeper understanding of faults, facilitating the implementation of preventive and recovery strategies. Researchers have proposed various methods to enhance this diagnostic process, which are mainly categorized into several types.

- Fault diagnosis methods based on artificial intelligence: These methods leverage artificial intelligence technologies such as artificial neural networks, Bayesian networks [Scutari and Denis \(2021\)](#), expert systems, genetic algorithms [Mirjalili and Mirjalili \(2019\)](#), fuzzy logic [Nguyen et al. \(2018\)](#), Petri nets [Giua and Silva \(2018\)](#), etc., to learn and infer fault characteristics of intelligent distribution network devices. They achieve fault detection, localization, and diagnosis. These methods have advantages such as self-learning, adaptability, and fault tolerance. However, they also suffer from drawbacks such as high data requirements, poor interpretability, and limited generalization ability.
- Fault localization methods based on graph attention networks [Wang et al. \(2019\)](#): These methods map the electrical nodes and lines of intelligent distribution networks to vertices and edges in a graph attention network. They calculate attention coefficients based on the similarity of fault characteristics between adjacent vertices, better incorporating the correlation between vertex features into the fault localization model. This improves the model's adaptability to topological changes. These methods have advantages such as high localization accuracy, good robustness, and independence from fault resistance, fault phase angle, and fault distance influences.
- Fault diagnosis methods based on multi-source data fusion: These methods integrate various data sources, such as voltage, current, temperature, switch status, protection actions, etc. Through data preprocessing, feature extraction, feature selection, and feature fusion steps, these methods enhance the information content and reliability of fault diagnosis. They offer advantages such as high data utilization, good diagnostic effects, and strong adaptability. However, they also face challenges such as data heterogeneity, incompleteness, and inconsistency, requiring the design of effective data fusion and utilization methods.

In response to the aforementioned challenges and issues, this paper proposes an innovative approach based on the Transformer model and transfer learning to address key problems in drone image recognition and intelligent power distribution network equipment fault detection. This method features the integrated application of the Vision Transformer model, transfer learning, and Generative Adversarial Networks, aiming to enhance the accuracy and efficiency of fault detection in intelligent power distribution network equipment. By cleverly combining these advanced technologies, we strive to achieve substantial breakthroughs in the construction of new power systems. Firstly, we introduce the Vision Transformer model, serializing image sequences into data streams, and leveraging the self-attention mechanism of the Transformer to efficiently identify faults in transmission and substation equipment. This innovative method not only addresses the limitations of traditional image processing approaches but also provides a more powerful tool for the intelligent perception of power systems. Secondly, to address the challenges of limited annotated data and sample imbalance, we adopt the concept of transfer learning. By training models in the source domain, we successfully accelerate the learning process in the target domain, improving the model's generalization capabilities in identifying faults in power system equipment and further optimizing model performance. To enhance

the model's robustness and generalization capabilities, we also introduce Generative Adversarial Networks. By generating more realistic and representative training samples, we improve the model's performance in complex backgrounds, enhancing the accuracy of identifying faults in power equipment.

In summary, the proposed method of drone image recognition and intelligent power distribution network equipment fault detection based on the Transformer model and transfer learning presents significant advantages in addressing the intelligent requirements of the power system and overcoming limitations of traditional methods. Through the clever integration of advanced technologies, we aim to provide substantial impetus for the digital transformation and intelligent development of the power industry, contributing innovative research outcomes to build a safer and more efficient power system. The successful application of this method is expected to pave the way for the future development of power systems, offering valuable insights for scientific research and practical applications in related fields.

The contributions of this paper can be summarized in the following three aspects:

1. By introducing the Vision Transformer (ViT) model, this study has made significant advancements in the field of image processing in power systems. ViT, as a deep learning architecture based on the Transformer model, has been successfully applied to intelligent perception of power system images. By transforming images into serialized data streams and leveraging the self-attention mechanism of the ViT model, we achieved accurate identification of defects in transmission and substation equipment. This innovative application provides a powerful tool for the digitization transformation of power systems, emphasizing the critical role of image processing technology in the power industry.
2. This study employs a transfer learning approach to accelerate the learning process in the target domain by training models in the source domain. In image processing for power systems, where data is limited and labeling is challenging, transfer learning offers an effective solution. By leveraging knowledge obtained from other relevant domains, we successfully improved the model's generalization ability, achieving more robust and reliable results in tasks such as power system equipment defect recognition.
3. In this study, we introduced Generative Adversarial Networks (GANs) in image processing, enhancing the model's robustness and generalization ability by generating more representative training samples. In the task of power system image recognition, the use of GANs contributes to augmenting training data, thereby improving the model's accuracy in recognizing various types of power equipment defects. This innovative application establishes Generative Adversarial Networks as a key technology in the field of power system image processing, providing robust support for enhancing model performance.

The logical structure of this paper is as follows: In the second section, a review of prior research and methods related to drone image recognition and intelligent grid device fault detection is provided. The strengths and weaknesses of existing methods and the challenges they face are analyzed, leading to the introduction of the

innovative points and solutions addressed in this study. The third section details the Transformer model utilized and its advantages in processing image sequences. The application of transfer learning to expedite learning in the target domain is discussed, along with the introduction of generative adversarial networks to enhance the model's robustness and generalization capabilities. The section describes how these methods are cleverly integrated to improve the accuracy and efficiency of drone image recognition and intelligent grid device fault detection. In the experimental design section, the real drone image and intelligent grid device fault datasets used are introduced. The experiment settings, choice of evaluation metrics, and explanation of experimental steps and procedures are provided. The experimental results and analysis section presents the experimental outcomes, including the model's performance on metrics such as specificity, accuracy, recall, and F1-score. Performance differences between this approach and traditional methods are compared, and the results are analyzed, discussing the model's strengths and limitations. Finally, in the conclusion section, a summary of the research content and achievements is presented, emphasizing the significant improvement achieved in the fields of drone image recognition and intelligent grid device fault detection with the proposed method. Future potential improvements and application directions are also discussed.

2 Related work

With the flourishing development of drone technology and artificial intelligence, the fields of drone image recognition and intelligent grid device fault detection have attracted significant attention [Azar et al. \(2021\)](#); [Ning et al. \(2024\)](#). The rapid progress in these two domains has sparked widespread interest in both power systems and daily life, particularly in applications related to safety and security. Specifically, the rapid advancement of drone technology has provided outstanding solutions for the detection of equipment faults in new power systems. Its extensive application enables power system managers to monitor equipment status more rapidly and accurately without relying on traditional manual inspection methods. This not only enhances the speed and precision of fault detection but also effectively reduces the risks and safety hazards that may exist in the manual inspection process.

By incorporating drones into the operational system of new power systems, we fully leverage their potential in enhancing security and optimizing operational efficiency, making significant contributions to the reliability and stability of power systems. Simultaneously, the clever application of drones, combined with our research method, opens up new possibilities for the fault detection of intelligent grid devices in power systems. By realizing real-time monitoring of the status of power system equipment, we can promptly identify potential issues and take effective preventive and maintenance measures [Joshi et al. \(2022\)](#). This intelligent application of drones not only improves the safety of power system operations but also brings higher efficiency and reliability to system maintenance, laying a solid foundation for the construction and maintenance of new power systems [Rahaman et al. \(2022\)](#).

In this context, the paper proposes a comprehensive approach based on the Transformer model and transfer learning, aiming to enhance the accuracy and efficiency of drone image recognition and

intelligent power distribution network equipment fault detection. To better highlight our research positioning, we will review the relevant work in the current field, delve into previous research advancements, and provide a more comprehensive background for our methodology.

Kumar et al. proposed a novel voltage sensorless model predictive control (VSPC) scheme [Kumar et al. \(2023\)](#). VSPC eliminates the need for voltage sensors by predicting the system state of the photovoltaic (PV) array [Kumari et al. \(2023\)](#). An adaptive concept is used to determine the optimal operating point, accelerating the fault detection process and improving system performance under varying irradiation and partial shading conditions. Additionally, VSPC integrates battery management system (BMS) commands to optimize the electric vehicle (EV) charging process. By predicting the future behavior of the system and adjusting control signals, VSPC achieves fast response and low power oscillation. Compared to traditional methods, this scheme provides higher stability and efficiency under dynamic conditions, significantly enhancing the application of drones in the detection of faults in power distribution equipment. In the study presented in [Lee et al. \(2018\)](#), a machine learning-based drone detection system was proposed. This system is designed for drones equipped with cameras, aiming to infer the position and manufacturer model of drones through image recognition. Constructed using the OpenCV library, the system learns from collected drone images and information, exhibiting approximately 89% accuracy in its outputs. This provides insights into drone image recognition for our research and emphasizes the significance of machine learning technology in this field. Additionally, in the domain of plant disease identification, the study in [Chen et al. \(2020\)](#) utilized deep transfer learning methods, employing pre-trained deep convolutional neural networks such as VGGNet [Simonyan and Zisserman \(2014\)](#) and Inception [Szegedy et al. \(2016\)](#) models. Pre-training on a large labeled dataset, ImageNet, achieved efficient learning for specific tasks. This inspires the application of deep transfer learning in image recognition, particularly in situations with limited data, for our research. The comprehensive review paper [Appiah et al. \(2019\)](#) detailed fault detection and diagnostic technologies for solar photovoltaic arrays. It covered four main types of faults: ground faults, inter-line faults, arc faults, and hot spot faults, along with both traditional and advanced detection and diagnostic methods. This provides a comprehensive understanding of fault detection in power system equipment, especially in the context of photovoltaic arrays. Literature [Abid et al. \(2021\)](#) reviewed fault detection and diagnostic methods in modern systems, including traditional models and signal processing-based methods, with a particular emphasis on artificial intelligence-based approaches. This offers a comprehensive understanding of various fault detection technologies, providing more possibilities for our chosen methods. In [Li et al. \(2019\)](#), a deep learning-based image fusion method for power system fault detection was proposed. Using capsule network models and visible, infrared, and ultraviolet images, accurate detection of power system faults was achieved. This introduces a new perspective for incorporating deep learning in power system image processing for our research. The paper [Mohammadi et al. \(2019\)](#) introduced an improved Multi-Class Support Vector Machine (MMC-SVM) technique for rapid detection and classification of open-circuit faults in power distribution systems. Efficient fault detection in

complex systems was achieved by utilizing the RMS voltage of the power grid. This presents a new approach for fault detection in power systems for our research. In the field of single-image super-resolution reconstruction, [Zhu et al. \(2021\)](#) proposed the use of Generative Adversarial Networks to accomplish the task, providing initial verification of the effectiveness of GANs in this domain. This work serves as a reference for employing GANs in image processing. Additionally, [Ning et al. \(2023\)](#) introduced a deep learning algorithm based on the long-tailed coverage function neural model, demonstrating its superiority in image classification tasks. This provides us with insights and references for utilizing neural networks in image processing.

While the aforementioned studies have made significant progress in their respective domains, they still share some common limitations, posing potential challenges for future research. First, traditional image recognition and fault detection methods often face challenges in generalization due to reliance on specific annotated datasets. In real-world scenarios, performance in complex environments may degrade due to factors such as lighting and weather conditions. Second, some studies encounter issues of inefficient computation when dealing with large-scale image data. Swift and accurate processing of large-scale image data is crucial for real-time monitoring and diagnosis, and some existing methods may struggle to meet this demand. Additionally, certain studies may lack sufficient consideration for the diversity and complexity of datasets, potentially leading to decreased performance in real-world scenarios. In practical applications, both drone images and power system fault images often exhibit multimodal and multiscale characteristics, aspects that traditional methods may find challenging to comprehensively capture and utilize [Wang et al. \(2024\)](#).

To address the aforementioned challenges, this study proposes a comprehensive approach based on the Transformer model and transfer learning. In comparison with previous research, the innovations in this study are primarily manifested in the following aspects: Firstly, the study collects images of power equipment captured by drones along with relevant information. Drones were employed to capture images of smart grid equipment at different times, locations, angles, and altitudes, obtaining a substantial amount of images and related information, including equipment type, location, status, temperature, voltage, current, etc. Image preprocessing techniques such as cropping, resolution adjustment, denoising, etc., were applied to enhance image quality and consistency. Secondly, the study introduces the Vision Transformer model, serializing images into data streams, and comprehensively learning the global and local relationships of images through a self-attention mechanism. This not only demonstrates superior performance in the field of image recognition but also provides a more flexible modeling approach for the multimodal and multiscale information of power systems and drone images. Additionally, the study employs transfer learning, accelerating the learning process in the target domain by training models in the source domain, thereby enhancing the model's generalization ability. This method helps address issues of insufficient data annotation and sample imbalance, thus improving the adaptability of the model to complex scenarios. Finally, the study introduces Generative Adversarial Networks (GANs), enhancing the model's performance in complex backgrounds by generating more representative training

samples. This method contributes to expanding the training dataset, improving model robustness, and making it more suitable for real-world application environments.

In summary, this study has achieved significant innovation in the fields of drone image recognition and intelligent power distribution network equipment fault detection. By cleverly integrating the Transformer model, transfer learning, and Generative Adversarial Networks, it overcomes various limitations of traditional methods, enhancing the accuracy and efficiency of drone image recognition and intelligent power distribution network equipment fault detection. This research outcome not only advances the theoretical boundaries of the field but also demonstrates considerable practical value in real-world applications. Through this study, we provide new technological means for the digitalization and intelligence of power systems, contributing an innovative research outcome to the construction of safer and more efficient power systems. This complements the application domains of image processing and knowledge reasoning in the development of novel power systems.

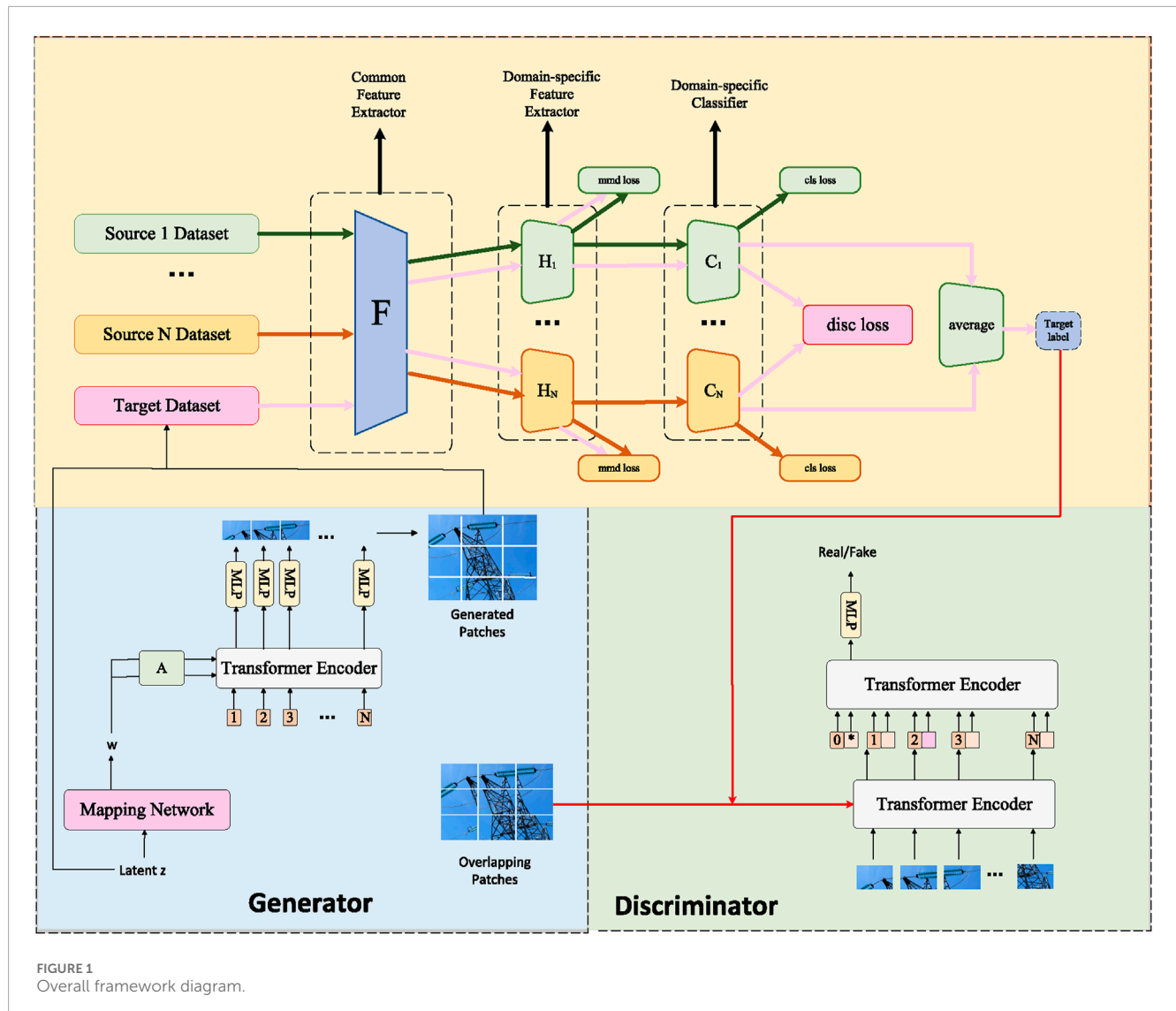
3 Methodology

In the methodology section of this study, we will provide a detailed explanation of the three key methods employed: the Vision Transformer model, Generative Adversarial Networks, and Transfer Learning. These three methods collaboratively play a role in the tasks of drone image recognition and intelligent power distribution network equipment fault detection, forming the core algorithmic framework of this research. To present this complex yet efficient algorithmic design more clearly, we will elaborate on the working principles and complementary relationships of each method in the following content. The overall algorithmic framework is illustrated in [Figure 1](#).

3.1 Vision transformer model

The Vision Transformer (ViT) is a model designed for image classification [Han et al. \(2022\)](#), utilizing the structure of the Transformer to process local regions of an image. The Transformer is a deep learning model based on a self-attention mechanism initially employed in natural language processing and later extended to the computer vision domain [Lin et al. \(2022\)](#). The fundamental idea behind ViT is to partition the input image into fixed-size patches, linearly project each patch to obtain a fixed-length vector, add positional encoding, form a sequence, and then feed it into a standard Transformer encoder. Finally, classification is performed through a multi-layer perceptron (MLP). The model architecture of ViT is illustrated in [Figure 2](#).

In our study, the input image has dimensions $H \times W \times C$, where H and W represent the height and width of the image, and C is the number of channels (RGB) in the image. The image is divided into N patches of size $P \times P \times C$, where $N = HW/P^2$. Each patch undergoes a linear transformation layer, resulting in a D -dimensional vector known as patch embedding. To retain the positional information of the patches in the image, a learnable position encoding is added, which is summed with the patch embedding to create an



$N \times D$ matrix. Additionally, for image classification, a special class embedding is added at the beginning of the sequence, serving as the label for classification. Therefore, the final input sequence has dimensions $(N + 1) \times D$.

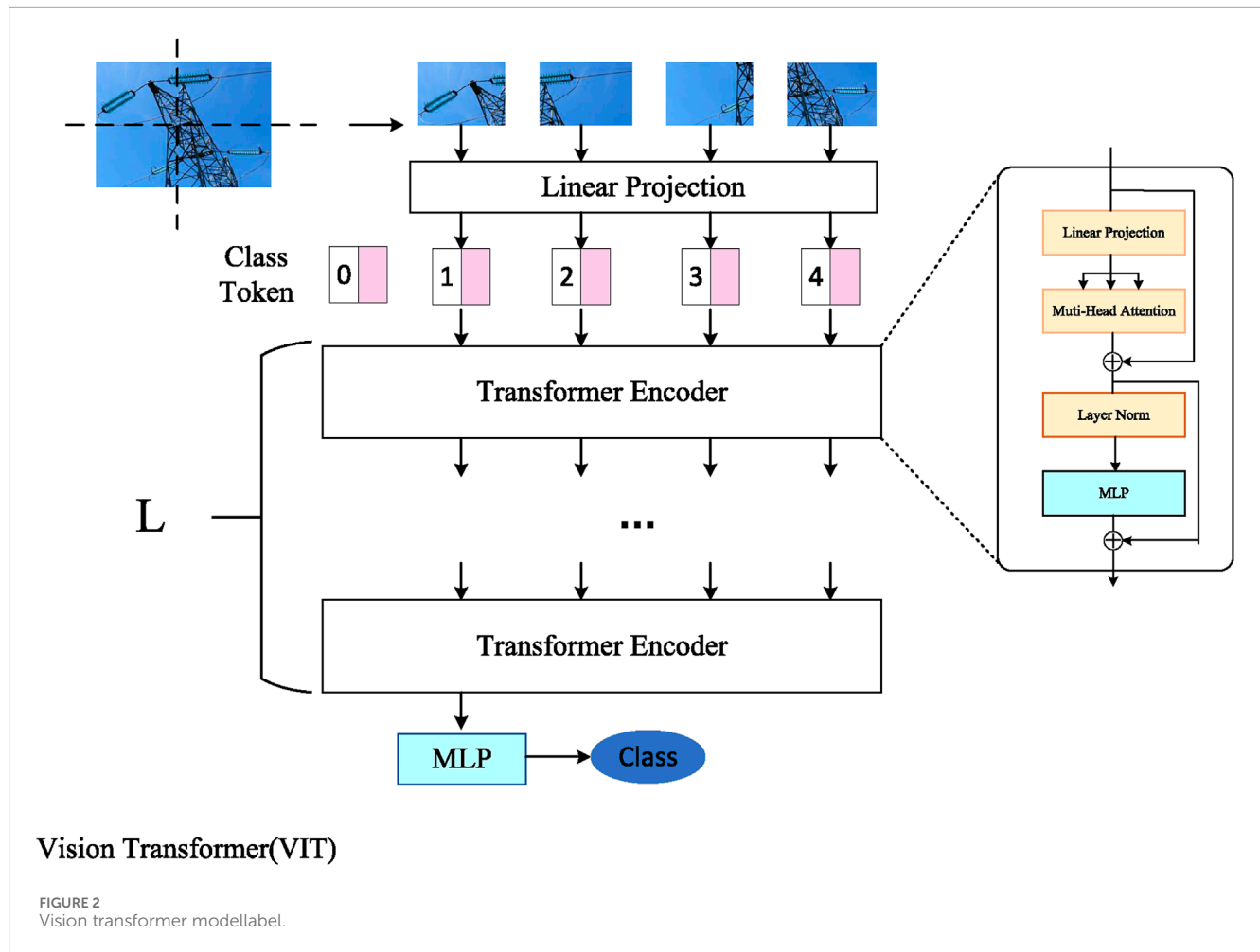
The input sequence passes through L Vision Transformer blocks (ViT blocks), each ViT block consisting of the following components:

- Layer Normalization (LN): Normalizes each vector in the input sequence, making its mean 0 and variance 1, which is beneficial for the convergence and generalization of the model.
- Multi-Head Self-Attention (MHSA) Voita et al. (2019): Peng et al. Ning et al. (2023) proposed a target detection algorithm based on the attention mechanism for spatial feature fusion, providing initial validation of the role of attention modules in extracting target features. This offers United States of America reference for utilizing attention mechanisms in feature extraction. Performs self-attention calculations on

each vector in the input sequence, meaning that, based on the correlation with other vectors, each vector is assigned a weight, and then a weighted sum is obtained, resulting in a new vector. Multi-head self-attention involves splitting the input sequence into multiple subsequences, performing self-attention calculations on each subsequence, concatenating the results, and then applying a linear transformation layer to obtain a matrix with dimensions $N \times D$. Multi-head self-attention can enhance the model's representational capacity and parallelism while reducing the number of parameters and computations. The calculation formula for multi-head self-attention is as follows:

$$\text{MHSA}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^0$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Given an input matrix X with dimensions (N, d_{model}) , where N is the sequence length and d_{model} is the dimensionality of the input vectors, the mechanism first applies three linear transformations to project X into query (Q), key (K), and value (V) matrices:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V$$

Here, W^Q , W^K , and W^V are trainable weight matrices of dimensions (d_{model}, d_k) and (d_{model}, d_v) . Each attention head i computes the scaled dot-product attention. W_i^Q , W_i^K , and W_i^V are the weight matrices for head i , and d_k is the dimensionality of the queries and keys, typically $d_k = d_v = \frac{d_{\text{model}}}{h}$ with h being the number of heads. The factor $\frac{1}{\sqrt{d_k}}$ stabilizes gradients by scaling the dot products. The outputs of the attention heads are concatenated, where $\text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)$ forms a matrix of dimensions $(N, h \cdot d_v)$, and W^O is a trainable weight matrix of dimensions $(h \cdot d_v, d_{\text{model}})$. This concatenation followed by the linear transformation results in the final output, effectively aggregating diverse features learned from multiple subspaces.

- **Residual Connection:** Add the output of the self-attention to the input, resulting in a matrix of dimensions $N \times D$, promoting model depth and stability.
- **Layer Normalization:** Normalize the output of the residual connection, obtaining a matrix of dimensions $N \times D$.
- **Multi-Layer Perceptron (MLP):** Apply two linear transformations to each vector in the normalized output, with an intermediate activation function (GELU), resulting in a matrix of dimensions $N \times D$. MLP enhances model nonlinearity and complexity, improving the model's fitting capability. The computation formula for MLP is as follows:

$$\text{MLP}(x) = xW_1 + b_1W_2 + b_2$$

Where x is the input vector, W_1 , W_2 , b_1 , b_2 are learnable weight matrices and bias vectors, and GELU is the Gaussian Error Linear Unit. The computation formula for GELU is as follows:

$$\text{GELU}(x) = x\Phi(x) = x\frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

Where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution, and $\text{erf}(x)$ is the error function.

ViT, as a new paradigm in image processing, provides us with powerful feature extraction tools for our research. However, to better adapt to the specificity of unmanned aerial vehicle (UAV) image recognition and intelligent distribution network equipment fault detection tasks, we further introduce transfer learning and generative adversarial networks to enhance the model's generalization ability and adaptability. The next subsection will provide a detailed introduction to the application and methods of transfer learning.

3.2 Transfer learning

Transfer learning is a machine learning method that leverages existing relevant knowledge to assist in learning new tasks, thereby improving learning efficiency and performance (Zhuang et al., 2020). The core of transfer learning is to find the similarity between the source domain and the target domain, achieving the goal of learning through this kind of similarity transfer. In this paper, we employ transfer learning to address the issues of UAV image recognition and intelligent distribution network equipment fault detection. Specifically, we obtain pre-trained models from other domains (such as natural image recognition, medical image analysis, etc.) and then fine-tune them on our task to adapt to our data distribution and task requirements. The general process of transfer learning is illustrated in Figure 3.

This involves the following steps:

1. Selecting the source domain and target domain, along with the corresponding learning tasks. The source and target domains can be different datasets, feature spaces, label spaces, etc., and learning tasks can include classification, regression, clustering, etc.
2. Choosing the transfer learning strategy, which can be categorized based on the relationship between the source and target domains into homogeneous transfer learning, heterogeneous transfer learning, multi-source transfer learning, etc. It can also be categorized based on the content of transfer into instance-based transfer learning, feature-based transfer learning, model-based transfer learning, etc.
3. Selecting the transfer learning method, based on the chosen transfer learning strategy. Different algorithms can be employed for transfer learning. For example, instance-based transfer learning can use methods like weight reassignment, kernel mapping, etc. Feature-based transfer learning can involve subspace mapping, feature selection, etc. Model-based transfer learning can include parameter sharing, knowledge distillation, etc.
4. Evaluating the effectiveness of transfer learning. Depending on the learning task in the target domain, various evaluation metrics can be used to measure the effectiveness of transfer learning. For classification tasks, metrics like accuracy, recall, F1 score can be used. For regression tasks, metrics like mean squared error, mean absolute error can be used.

In this paper, we adopt a model-based transfer learning approach. Specifically, we use a pre-trained Vision Transformer model as the source domain model and fine-tune it on the target

domain data to adapt to our task. Our transfer learning approach can be expressed using the following formula:

$$f_t(x) = g_t(h_t(x))$$

Where $f_t(x)$ is the prediction function for the target domain, $g_t(\cdot)$ is the classifier for the target domain, $h_t(\cdot)$ is the feature extractor for the target domain, and x is the input image. Our objective is to optimize the parameters of $f_t(x)$ through transfer learning so that it achieves the best performance on the target domain data. To achieve this goal, we use the following loss function for optimization:

$$L_t = -\frac{1}{N_t} \sum_{i=1}^{N_t} l_i$$

$$l_i = y_i \log f_t(x_i)$$

Where N_t is the amount of data in the target domain, l_i is the loss for the i th sample, y_i is the true label in the target domain, and $\log f_t(x_i)$ is the predicted probability in the target domain. We use stochastic gradient descent (SGD) (Haji and Abdulazeez, 2021) or other optimization algorithms to minimize this loss function, thereby achieving transfer learning.

By introducing transfer learning, our model can better adapt to the characteristics of the target domain, improving the generalization performance in the tasks of unmanned aerial vehicle (UAV) image recognition and intelligent distribution network equipment fault detection. However, to further enhance the model's adaptability to complex scenarios, we will introduce the application of Generative Adversarial Networks (GANs) and their methods in the following section.

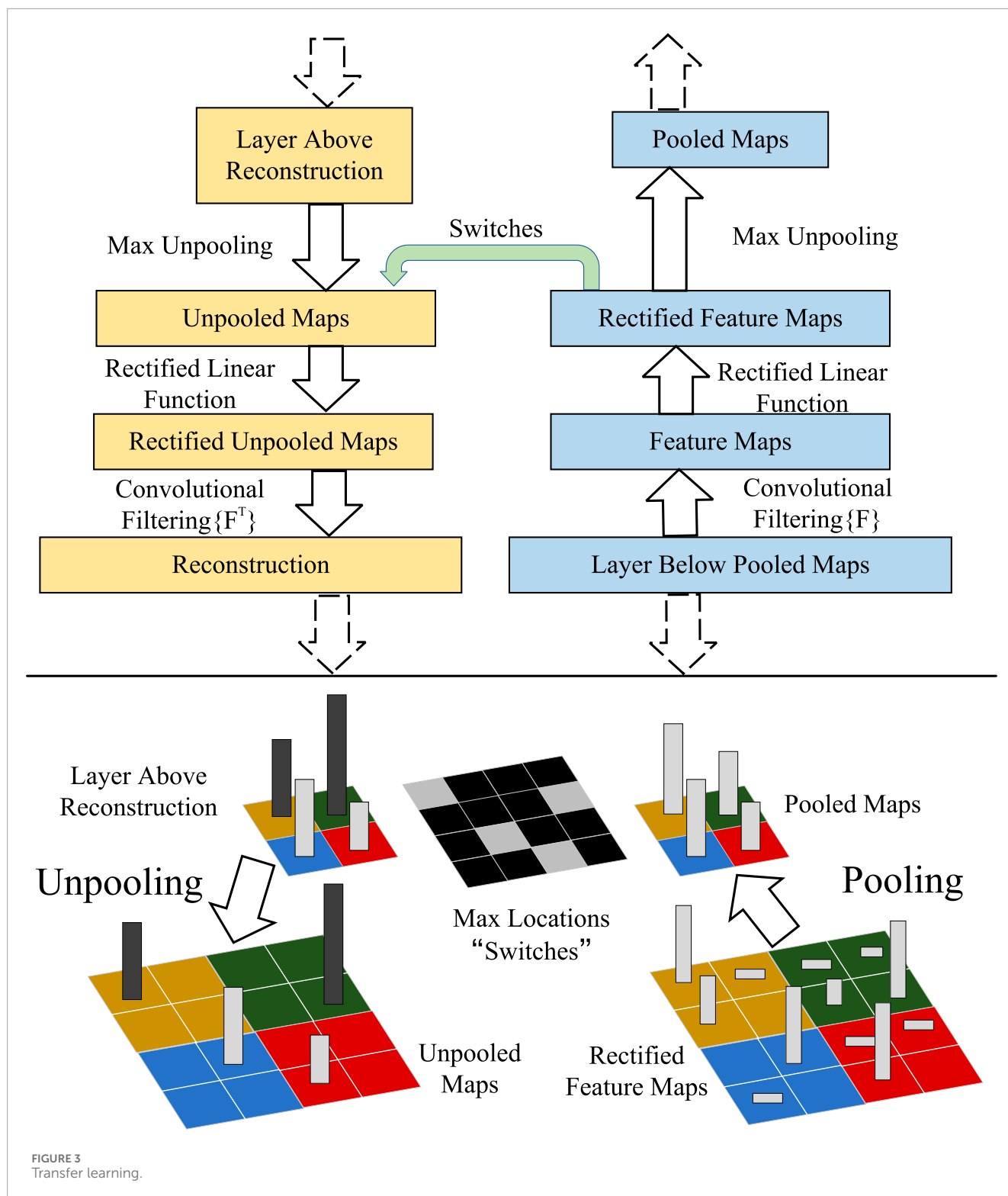
3.3 Generative adversarial networks

Generative Adversarial Network (GAN) is an unsupervised learning method consisting of two neural networks: a Generator and a Discriminator (Creswell et al., 2018). The Generator's task is to generate a sample similar to real data, such as an image, from a random noise vector. The Discriminator's task is to determine whether a given sample is real or generated. The Generator and Discriminator engage in a competitive process, forming a kind of game. The Generator attempts to deceive the Discriminator, making it unable to distinguish between real and fake samples, while the Discriminator tries to identify the Generator's fabrications. Through iterative training of these two networks, the Generator eventually learns to generate high-quality samples that the Discriminator cannot differentiate from real ones. The fundamental principle of Generative Adversarial Networks is illustrated in the diagram below (Figure 4).

The objective function of a Generative Adversarial Network can be expressed using the following formula:

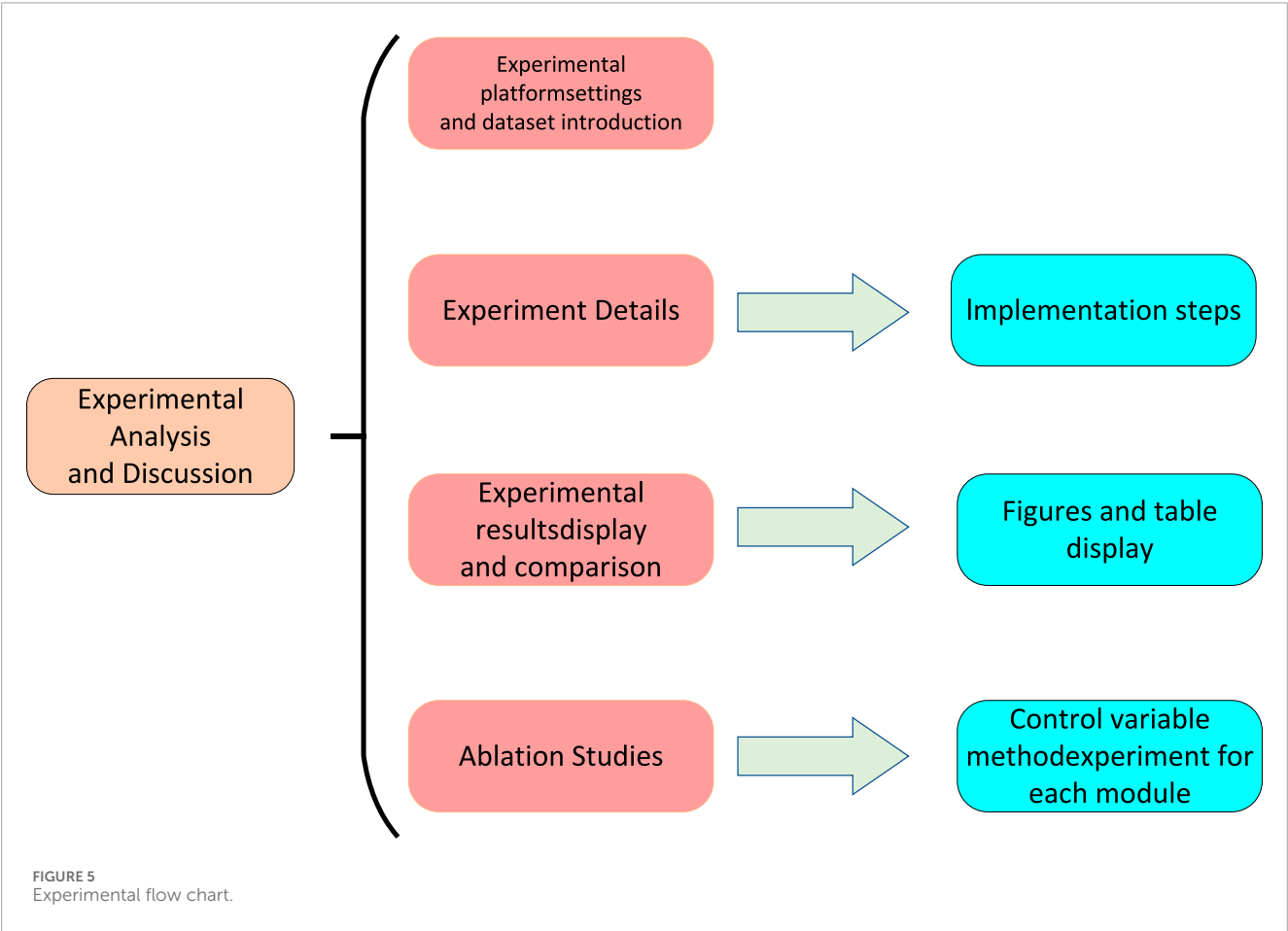
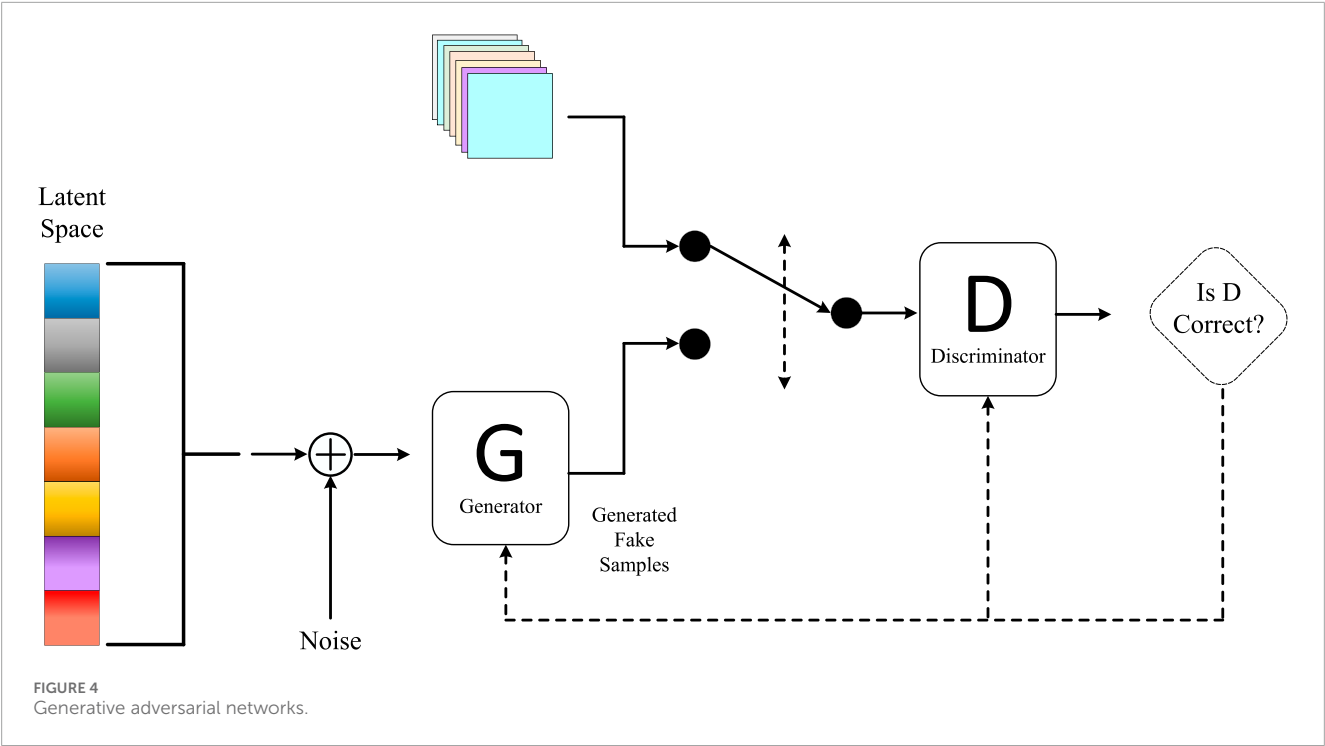
$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))]$$

Where $V(D, G)$ is the value function for the discriminator and generator, $D(x)$ is the output probability of the discriminator for a real sample x , $G(z)$ is the output sample of the generator



for a random noise vector z , $D(G(z))$ is the output probability of the discriminator for the generated sample $G(z)$, E is the expectation symbol, $p_{\text{data}}(x)$ is the distribution of real data, and $p_z(z)$ is the distribution of noise vectors. This objective function indicates that the discriminator aims to maximize the difference between the probabilities of real and generated samples, while

the generator aims to minimize this difference. To solve this optimization problem, a common strategy is to alternately update the discriminator D (keeping G fixed) to maximize $V(D, G)$ and update the generator G (keeping D fixed) to minimize $V(D, G)$. This process can be implemented using the following gradient descent algorithm:



$$\theta_d \leftarrow \theta_d + \alpha \nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))]$$

$$\theta_g \leftarrow \theta_g - \alpha \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)})))$$

Where θ_d and θ_g represent the parameters of the discriminator and generator, α is the learning rate, m is the batch size, $x^{(i)}$ and $z^{(i)}$ represent the i -th real data sample and noise vector, respectively.

In this paper, we employ the Generative Adversarial Network (GAN) method to enhance our transfer learning approach, aiming to improve the robustness and generalization capability of the model. Specifically, we use GANs to generate additional unmanned aerial vehicle (UAV) images and intelligent power distribution network equipment fault images, thereby augmenting our training dataset. To achieve this goal, we utilize the following GAN loss function:

$$L_{GAN}(G, D) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \\ + \lambda E_{x \sim p_{data}(x)} [(D(x) - 1)^2]$$

Where λ is a regularization coefficient used to penalize the discriminator for misjudging real data, enhancing the discriminator's discriminatory ability, and preventing the generator from converging too early to local optima.

By introducing the Generative Adversarial Network, our model is capable of generating samples in a more realistic and diverse manner, thereby improving the data utilization efficiency in the intelligent power distribution network equipment fault detection task. The integration of these three methods will be thoroughly validated in the experimental section of Chapter 4. Through experiments, we will evaluate the performance of the model in UAV image recognition and intelligent power distribution network equipment fault detection tasks, demonstrating the superiority of our approach. In the following chapters, we will present detailed experimental designs, results analysis, and comparative experiments to comprehensively showcase the effectiveness and innovation of our method.

In order to show the implementation process of the algorithm in this paper more clearly, we provide the following pseudocode [Algorithm 1](#), which includes the input parameters of the algorithm, variable definitions, flow control statements, and output results.

4 Experiment

As we delve into the detailed explanation of our method, this chapter will showcase empirical experiments conducted to validate our proposed approach. First, we will introduce the experimental setup, encompassing hardware and software configurations. Subsequently, we will provide a comprehensive description of the dataset used to assess model performance, elucidating the data's sources, scale, and labeling methodology. Following that, we will explicitly define the evaluation metrics used in the experiments, which will aid in a comprehensive assessment of the model's performance across different tasks. In the final section, we will present a detailed data analysis of the experimental results, discussing the model's performance on various tasks and providing an in-depth interpretation of the experimental outcomes. Through

Require: Source domain dataset D_s , target domain dataset D_t

```

1: Initialize Vision Transformer model  $M_{ViT}$ ,
   Transfer Learning model  $M_{TL}$ , GANs model  $M_{GAN}$ 
2: Initialize learning rates  $\eta_{ViT}, \eta_{TL}, \eta_{GAN}$ 
3: Initialize GANs hyperparameters  $\lambda, \alpha, \beta$ 
4: for each training epoch do
5:   for each mini-batch  $(x_i^s, y_i^s)$  in  $D_s$  do
6:     Update Vision Transformer model parameters
       using cross-entropy loss:

$$\mathcal{L}_{ViT} = -\frac{1}{N} \sum_{i=1}^N (y_i^s \log(P(M_{ViT}(x_i^s))))$$

7:   end for
8:   for each mini-batch  $(x_i^t)$  in  $D_t$  do
9:     Update Transfer Learning model parameters
       using domain adaptation loss:

$$\mathcal{L}_{TL} = \mathcal{L}_s + \lambda \cdot \mathcal{L}_{DA}$$

10:  end for
11:  for each mini-batch  $(x_i^s, y_i^s)$  in  $D_s$  do
12:    Generate synthetic samples using GANs:

$$x_i^{syn} = M_{GAN}(x_i^s)$$

13:    Update GANs model parameters using
       adversarial loss:

$$\mathcal{L}_{GAN} = \alpha \cdot \mathcal{L}_G + \beta \cdot \mathcal{L}_D$$

14:  end for
15: end for
16: Evaluation Phase:
17: for each mini-batch  $(x_i^t)$  in  $D_t$  do
18:   Use  $M_{ViT}$  and  $M_{TL}$  for target domain prediction
19:   Calculate Recall, Precision, and other
       evaluation metrics
20: end for

```

Algorithm 1. Integrated Model Training.

the content of this chapter, readers will gain a comprehensive and profound understanding of the effectiveness of our proposed method in practical applications. The Experimental flow chart is illustrated in [Figure 5](#) below.

4.1 Experimental environment

4.1.1 Hardware environment

This experiment utilized a high-performance computing server for hardware environment configuration. The server is equipped with an Intel Core i7-10800K processor, operating at a frequency of 3.70GHz, and boasts a memory capacity of 128GB RAM. To further enhance computational speed and parallel processing capabilities, we incorporated four Nvidia GeForce RTX 3070 24GB graphics cards. This powerful hardware configuration not only ensures the efficiency of experimental computations but also provides ample computational resources for the training and inference of deep learning tasks, contributing

to the accuracy and efficiency of the experiment. Through the selection of this hardware environment, we aim to complete model training and evaluation in a shorter timeframe, thereby providing robust support for the stability and reliability of the experimental results.

4.1.2 Software environment

In this study, we chose Python and PyTorch as the primary software environment for implementing the method of unmanned aerial vehicle (UAV) image recognition and intelligent fault detection of power grid devices. Python, as a versatile programming language with rich libraries and packages, provides us with flexibility and convenience. Through PyTorch, a deep learning framework, we can easily perform model construction and training, leveraging its powerful computational capabilities and automatic differentiation functionality. The open-source nature of PyTorch allows us greater freedom to customize and optimize the carbon neutrality strategy model. Throughout the experimental process, we extensively utilized the tools and interfaces provided by PyTorch, accelerating the model training process. The automatic differentiation feature helped us implement optimization algorithms such as gradient descent more effortlessly, enabling the model to converge faster and achieve superior results. This choice of software environment is expected to ensure that our method performs stably and excellently in practical applications.

4.2 Experimental data

4.2.1 ICS-SGAD dataset

The ICS-SGAD [Noda et al. \(2023\)](#) is a dataset containing intelligent grid communication, primarily involving two protocols: IEC 60870-104 (IEC 104) and IEC 61850 (MMS). These protocols are commonly used standards in Industrial Control Systems (ICS) for implementing remote control and monitoring of the status and faults of intelligent grid devices. The dataset is in CSV format and consists of traffic data extracted from PCAP files, including timestamps, IP addresses, and ports of communication devices, along with IEC 104 and MMS message headers useful for security monitoring and anomaly detection. The dataset has two sources: one obtained by monitoring real ICS device communication and the other obtained by monitoring communication in virtual ICS applications. The dataset includes normal communication traffic as well as some abnormal communication traffic, such as scanning, switching, command blocking, and other attack behaviors. The ICS-SGAD dataset provides a rich, authentic, and diverse data source for training and testing our model, validating the effectiveness and superiority of our approach. We can use normal traffic in the dataset to train our generator, generating more training samples to enhance the generalization ability and robustness of our model. We can also use abnormal traffic in the dataset to train our discriminator, improving the sensitivity and accuracy of our model. Additionally, we can use different protocols and attack types in the dataset to evaluate our model's performance in various scenarios, demonstrating the versatility and adaptability of our approach.

4.2.2 AKNN-SGFD dataset

The AKNN-SGFD dataset [Barta et al. \(2015\)](#) is designed for intelligent grid fault detection and classification, generated based on an improved K-nearest neighbors algorithm (AKNN). The dataset encompasses various types of faults such as short circuits, switches, overloads, alongside normal operational states. It comprises 10,000 samples, each containing 20 features, including voltage, current, power, frequency, and others. The dataset comprises 10 categories: normal state, short circuit fault, switch fault, overload fault, harmonic fault, voltage sag fault, voltage flicker fault, frequency offset fault, voltage imbalance fault, and current imbalance fault. The sample counts vary across categories, resulting in data imbalance. Each fault is labeled to indicate the type and location of the fault. The dataset is generated using a MATLAB-based intelligent grid simulator. Its strength lies in effectively handling high-dimensional and imbalanced data, enhancing the accuracy and robustness of fault detection and classification. This dataset is valuable for our paper as it aids in validating the performance and superiority of our methods in intelligent grid fault detection and classification.

4.2.3 FFC-SG dataset

The FFC-SG dataset [Nagy et al. \(2016\)](#) is designed for intelligent grid fault and failure classification, generated based on a rapid face classification (FFC) method. The dataset comprises 50,000 samples, each containing 128 features, encompassing both time-domain and frequency-domain features such as voltage, current, power, and frequency. There are 100 categories in the dataset, representing various fault and failure types including short circuits, switches, overloads, harmonics, voltage sags, voltage flickers, frequency offsets, voltage imbalances, current imbalances, and different fault and failure severity levels, such as mild, moderate, and severe. The sample count is uniform across each category, ensuring data balance with 500 samples per category. Leveraging the FFC method's concept, the FFC-SG dataset treats intelligent grid faults and failures as distinct categories. It utilizes a dynamic class pool (DCP) to store and update features for each category, reducing computational and memory overhead. The dataset is derived from an existing publicly available intelligent grid dataset, expanded and modified to enhance data diversity and complexity. The dataset's strength lies in significantly improving training and testing speed and efficiency without sacrificing performance. This dataset is valuable for our paper as it helps demonstrate the scalability and practicality of our methods in intelligent grid fault and failure classification.

4.2.4 OWA-SGFD dataset

The OWA-SGFD dataset [Neeraj and Behera \(2022\)](#) is designed for intelligent grid fault diagnosis and is generated based on an Ordered Weighted Averaging (OWA) fuzzy rough set method. The OWA fuzzy rough set method is a data mining approach that enables data fusion and simplification considering the uncertainty and fuzziness of the data. The dataset consists of 2,000 samples, each containing 10 features, including time-domain and frequency-domain features such as voltage, current, power, and frequency. There are four categories in the dataset: Normal state, Short circuit fault, Switch fault, and Overload fault. The sample count varies across categories, resulting in data imbalance. The OWA-SGFD dataset utilizes the OWA fuzzy rough set method to fuse and

TABLE 1 Data partitioning for different datasets.

Dataset	Training set	Test set	Validation set
ICS-SGAD	41,148	11,756	5,878
AKNN-SGFD	31,500	9,000	4,500
FFC-SG	52,500	15,000	7,500
OWA-SGFD	45,500	13,000	6,500

compress data from different sensors and devices, enhancing data quality and usability. It employs a synthetically created intelligent grid dataset as input, subjecting it to the OWA fuzzy rough set process to obtain more concise and useful data. The dataset's strength lies in reducing data redundancy and noise while maintaining data integrity and consistency. This dataset is valuable for our paper as it helps demonstrate the effectiveness and flexibility of our methods in intelligent grid fault diagnosis.

- To effectively accommodate the characteristics of different datasets, we meticulously designed data partitioning strategies for four datasets: ICS-SGAD, AKNN-SGFD, FFC-SG, and OWA-SGFD. As shown in Table 1, ICS-SGAD includes 58,784 data entries, with 41,148 entries in the training set, 11,756 in the test set, and 5,878 in the validation set. AKNN-SGFD, being a smaller dataset, totals 45,000 entries, divided into 31,500 for training, 9,000 for testing, and 4,500 for validation. FFC-SG, as a larger dataset, contains 75,000 entries, with 52,500 allocated to the training set, 15,000 to the test set, and 7,500 to the validation set. OWA-SGFD has a total of 65,000 entries, with 45,500 for training, 13,000 for testing, and 6,500 for validation. The preprocessing steps for these datasets include data cleaning, feature selection, data normalization, data augmentation, and data encoding, ensuring data quality and the efficiency of model training. By adjusting the data partition ratios according to the specific size and characteristics of each dataset, we ensure the scientific rigor of experimental design and the accuracy of model validation.

4.3 Evaluation index

When evaluating the performance of our research methodology in tasks related to unmanned aerial vehicle (UAV) image recognition and intelligent power distribution network equipment fault detection, we employed a set of key evaluation metrics covering various aspects of the model's performance. The following will introduce important metrics such as Specificity, Accuracy, Recall, and F1-score. Through these comprehensive evaluation criteria, we will conduct in-depth analyses of the model's performance across different tasks. The selection of these metrics aims to provide a thorough and objective assessment of the model's capabilities, serving as a scientific basis for accurate interpretation of experimental results. Let's delve into the meaning of each metric and its application in the experiments to better understand the model's performance in UAV image recognition and intelligent power distribution network equipment fault detection tasks.

4.3.1 Specificity

Specificity assesses the accuracy of the model in predicting negative cases, with a particular focus on the correct identification of images without faults and normal power distribution network devices. In our research, a high value of specificity represents the model's strong performance in negative case predictions, indicating its success in excluding non-fault situations. The formula for calculating specificity is as follows:

$$\text{Specificity (\%)} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \times 100\%$$

Where, True Negatives (TN): Represents the number of samples correctly predicted as negative cases, i.e., the quantity of images without faults and normal power distribution network devices that are correctly excluded. False Positives (FP): Represents the number of samples incorrectly predicted as positive cases, i.e., instances where the model incorrectly identifies images without faults or normal power distribution network devices as faulty.

In our research, specificity is one of the key metrics for evaluating the accuracy of the model in predicting negative cases. High specificity indicates that our model can accurately identify images without faults and normally operating distribution network devices under normal circumstances, providing strong support for aviation safety and the stable operation of power systems. Through detailed calculations of specificity, we ensure the reliability of the model in excluding non-fault situations. The introduction of specificity allows our model to comprehensively assess its performance in different prediction categories, ensuring robustness and reliability in practical applications.

4.3.2 Accuracy

In our study, accuracy is a crucial evaluation metric used to measure the overall performance of the model in unmanned aerial vehicle (UAV) image recognition and intelligent power distribution network device fault detection tasks. Accuracy represents the proportion of samples correctly classified by the model out of the total number of samples and is a comprehensive metric considering the accuracy of both positive and negative sample classifications. The formula for accuracy is as follows:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}} \times 100\%$$

Where, True Positives (TP): Represents the number of samples correctly identified as positive (faulty intelligent power distribution network devices or UAV images). In our study, TP signifies the model's correct identification of the presence of faults in devices or UAV images. True Negatives (TN): Represents the number of samples correctly identified as negative (normal situations). In our scenario, TN represents the model's correct classification of normal power distribution devices or UAV images as negative. Total Samples: Represents the total number of samples, i.e., the total number of samples evaluated by the model.

Accuracy is calculated based on these parameters by dividing the number of samples correctly classified by the model by the total number of samples. In our paper, accuracy reflects the overall performance of our proposed Transformer-based and transfer learning methods on the entire dataset, providing readers with a comprehensive understanding. High accuracy indicates satisfactory

results in classifying both positive and negative samples, showcasing the practical potential for UAV image recognition and intelligent power distribution network device fault detection.

4.3.3 Recall

Recall, in our study, is a crucial evaluation metric used to measure the model's capability in capturing true positives in UAV image recognition and intelligent power distribution network device fault detection tasks. Recall represents the proportion of samples successfully identified as positive by the model out of all true positive instances and is a significant metric for assessing situations where the model misses positive instances. The formula for recall is as follows:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \times 100\%$$

Where True Positives (TP) represents the number of samples correctly identified as positive (UAV images or intelligent power distribution network device faults) by the model. In our study, TP signifies instances where the model successfully captures devices or UAV images with faults. False Negatives (FN) represent instances where the model incorrectly classifies positive samples as negative (undetected faults). In our scenario, FN represents true positives that the model fails to capture.

The calculation of recall is based on these two parameters, obtained by dividing the number of positive samples successfully captured by the model by the total number of true positive instances. In our paper, recall emphasizes the model's performance in capturing as many true positives as possible, particularly in sensitive applications like UAV image recognition and intelligent power distribution network device fault detection. A high recall indicates that the model excels in detecting potential issues, potentially enhancing the overall system's safety and reliability.

4.3.4 F1-score

In our study, the F1-score is a comprehensive evaluation metric that considers both precision and recall, aiming to balance the model's performance on positive and negative samples. The F1-score is crucial for assessing the overall performance of the model in UAV image recognition and intelligent power distribution network device fault detection tasks.

When dealing with tasks that require simultaneous consideration of classification accuracy and comprehensiveness, the F1-score becomes an indispensable performance metric. Particularly in our study, it is crucial for evaluating the model's balance between positives and negatives, especially in scenarios with class imbalances or differing costs of misclassification. Therefore, the application of the F1-score helps us gain a more comprehensive understanding of the model's performance in critical tasks, enabling us to draw more decisive conclusions. The formula for calculating the F1-score is as follows:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%$$

In this context, the parameters have the following meanings: Precision: Represents the proportion of samples correctly predicted as positive by the model among all samples predicted as positive. In our study, this is the ratio of the number of samples correctly classified as positive by the model based on input data to

the total number of samples predicted as positive. Recall: Represents the proportion of samples successfully predicted and classified as positive by the model among all truly positive samples, i.e., recall. As introduced earlier.

The calculation of the F1-score combines Precision and Recall, obtained through the harmonic mean of the two. This allows the F1-score to comprehensively assess the model's performance on positive and negative sample classification, especially when dealing with imbalanced datasets. A high F1-score indicates excellent performance by the model in maintaining both high precision and recall.

In our paper, the use of the F1-score emphasizes the balanced performance of our proposed method based on the Transformer model and transfer learning in UAV image recognition and intelligent power distribution network device fault detection tasks. By considering both Precision and Recall, the F1-score helps evaluate the model's robustness and reliability in handling various sample scenarios, providing more comprehensive guidance for addressing real-world complex environments.

4.4 Experimental details

In this experiment, we performed a series of hyperparameter settings to optimize the model's performance. Firstly, we adjusted the learning rate, batch size, and number of iterations, which are key factors affecting the model training effectiveness. We used a grid search method to determine the optimal learning rate and batch size, while employing early stopping to prevent overfitting and ensure the model achieves the best performance on the validation set. Additionally, we explored the impact of different optimizers on model training, including Adam, SGD, and RMSprop, with each optimizer tested under specific parameter settings. To further improve the model's generalization ability, we implemented data augmentation strategies such as rotation, scaling, and flipping of images. Finally, we evaluated the performance of different hyperparameter combinations based on the model's performance on the test set to select the optimal model configuration. The specific hyperparameter settings are shown in [Table 2](#).

4.5 Experimental comparison and analysis

Our experiments aim to validate the superiority of our approach in handling real UAV images and intelligent power distribution network fault datasets. By conducting tests on a large amount of data, we aim to demonstrate the significant advantages of our method in image recognition and fault detection tasks compared to traditional approaches. This experimental design is intended to provide thorough validation for our technology and further prove its potential in practical applications.

Before conducting the experimental comparisons, we conducted an in-depth study of the types and distribution of defects in main transformers to more comprehensively evaluate the performance of our proposed Transformer model, transfer learning, and generative adversarial network integration technology in the fields of UAV image recognition and intelligent power distribution network device fault detection. To achieve this, we introduced a crucial visualization

TABLE 2 Hyperparameter settings.

Hyperparameter	Setting
Learning Rate (Generator)	0.0002
Learning Rate (Discriminator)	0.0005
Optimizer	Adam, $\beta_1 = 0.5$, $\beta_2 = 0.999$
Batch Size	128
Epochs	300
Latent Space Dimension	100
Generator Layers	15
Discriminator Layers	12
Dropout Rate	0.3
Image Size	256 * 256
Label Smoothing	0.9

element, namely, the statistical chart of the number of main transformer defect types, as illustrated in Figure 6 below:

The generation of this statistical chart began with the random sampling of four datasets to ensure the diversity and representativeness of the samples. Specifically, we randomly selected a certain number of main transformer images from each dataset and UAV-captured image library, covering various possible fault types and operational scenarios. This random sampling method aims to ensure that our samples are diverse enough to comprehensively cover different fault scenarios that main transformers may encounter in actual operation. Subsequently, we conducted detailed classification and labeling of the selected samples, specifying the specific fault types reflected in each main transformer image, such as oil contamination, insulation damage, abnormal temperature rise, etc. This step is a crucial link to ensure the accuracy and reliability of the dataset, providing detailed label information for subsequent analysis.

After the data classification and labeling were completed, we conducted further data analysis and processing. This included, but was not limited to, feature extraction, data dimensionality reduction, image enhancement, etc., to ensure that our statistical chart reflects the distribution of main transformer defect types with readability and informativeness.

Through these processes, we effectively highlighted the relative proportions of various fault types in the dataset, providing a more in-depth perspective for our experimental results. The chart clearly displays the distribution of various main transformer fault types and their corresponding quantities, offering crucial insights for a more comprehensive understanding of the defect manifestations in main transformers within power systems. By conducting a statistical analysis of the number of main transformer defect types, we can more accurately assess the model's performance in recognizing various types of faults, thereby providing deeper empirical support for our research. The introduction of this statistical chart not only

makes the experimental results more intuitive but also provides more detailed empirical data, further solidifying our innovative research in the field of power system image processing.

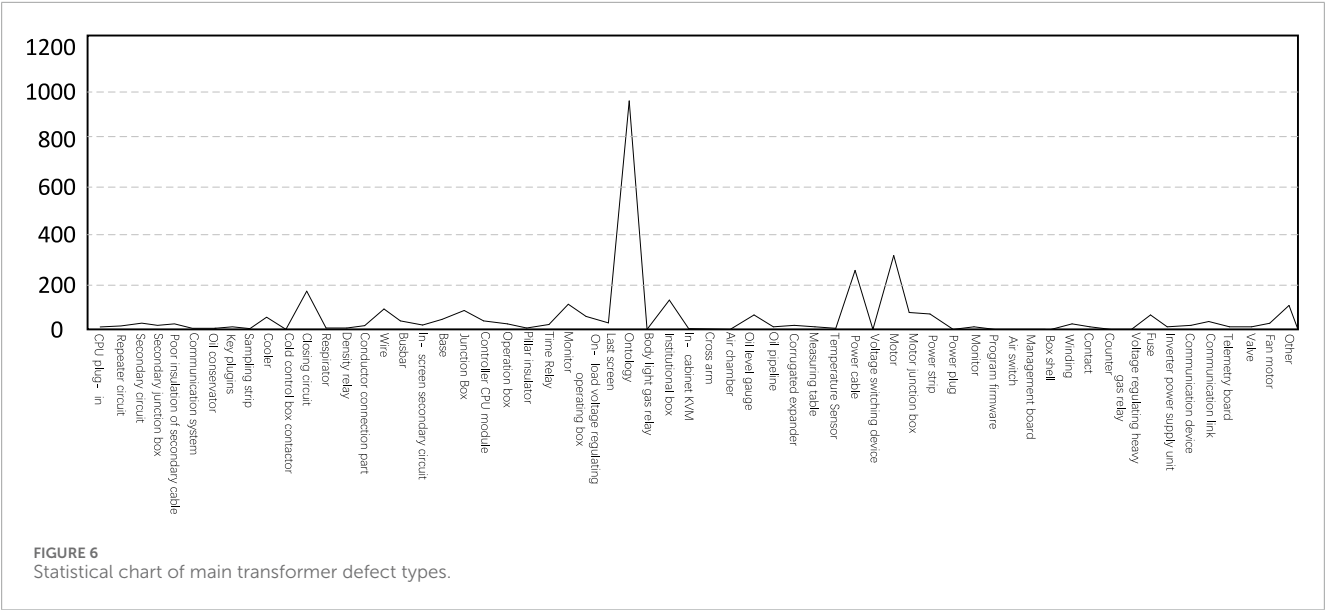
In further research analysis, we introduced a Transformer Fault Types chart, as shown in Figure 7. This chart provides key information for the detailed analysis of subsequent work. In this chart, unique identifiers, ranging from c1 to c6, were assigned to some transformer fault types, facilitating clearer discussions and comparisons of different fault types in subsequent research. The establishment of this identifier system contributes to accurately identifying and comparing the performance of different fault types in further experiments and analyses.

Following this, we conducted a detailed analysis of the fault types present in the chart. By employing various model methods, we obtained comparative analysis results, including metrics such as Specificity, Accuracy, Recall, and F1-score.

Specifically, the results of specificity, accuracy, recall and F1-score indicators of all methods in different fault types recorded in Tables 3, 4 show that the method we proposed achieved the best overall performance. Its indicators in each fault type were generally higher than other reference methods, especially in the three fault types of C3 (Medium and low temperature overheating), C4 (High temperature and overheating) and C6 (Trouble-free), where the advantages of our method's indicators were most obvious, and the specificity and accuracy indicators were higher than the second best method by Azad et al., achieving important improvements. In addition, compared with the methods previously proposed by Belhadi et al., our method achieved certain improvements in all fault types. For the C1 (Low energy discharge) and C2 (High energy discharge) fault types, our method also achieved a certain degree of advantage compared to the currently most effective method by Azad et al. Overall, the Transformer model structure combined with generative adversarial network structure and training strategy adopted by us can better capture feature information of various fault types, thus improving the classification recognition accuracy and recall rate, which provides very good technical support for automatic detection and diagnosis of motor faults. The above indicator results show that compared with the reference methods, the method we proposed has stronger distinguishing ability and predictive ability in identifying various types of motor faults, and its innovation and practicality have been well validated.

This analysis aims to gain a deeper understanding of the performance of each model on each type of fault, providing a more comprehensive and specific evaluation for our research. Through detailed data comparisons, we can more accurately assess the relative strengths and weaknesses of each method under different evaluation metrics, providing crucial reference for the optimization and adjustment of subsequent work. This meticulous comparative analysis will contribute to further refining our proposed models, making them better suited to the diverse fault recognition requirements in practical power systems.

Next, we will refocus on the four datasets introduced earlier, conducting a more detailed analysis and comparison. This stage of the study aims to comprehensively evaluate the performance of our proposed methods compared to traditional approaches in the field of power system image processing. We will primarily assess Specificity, Accuracy, Recall, and F1-score, while also examining key performance parameters such as training time, inference time,



Serial number	Fault type	Serial number	Fault type
C ₁	Low energy discharge	C ₄	High temperature and overheating
C ₂	High energy discharge	C ₅	Partial Discharge
C ₃	Medium and low temperature overheating	C ₆	Trouble-free

FIGURE 7
Transformer fault types.

and model parameters. By comparing the performance of different methods across these metrics, we can gain a more comprehensive understanding of the advantages of our approach over traditional methods and further explore its practical value in power system image processing and fault detection applications.

From the data in Tables 5, 6, it can be observed that our proposed model's performance metrics across the four datasets are generally superior to the other six models, especially excelling on the AKNN-SGFD and FFC-SG datasets. Specifically, our model on the FFC-SG dataset outperforms Azad et al.'s model by 2.6%–3.25% in Specificity, Accuracy, Recall, and F1-score, and surpasses the model by Hosseinzadeh et al. by 3.73%–4.28%. On the OWA-SGFD dataset, our model achieves a specificity metric 2.09% higher than Azad's and surpasses Belhadi et al.'s model by 8.48%. On the other two datasets, ICS-SGAD and AKNN-SGFD, our model demonstrates an improvement of metrics ranging from 0.91% to 1.83% compared to Azad's model and a higher improvement

of 1.45%–2.5% compared to Hosseinzadeh's model. Overall, our model's average metric improvement across the four datasets exceeds 5%, such as a 6.25%–7.5% improvement compared to Belhadi's model on the ICS-SGAD dataset and a 7.93%–8.49% improvement on the AKNN-SGFD dataset. This strongly indicates that the technical approaches employed in our proposed model effectively enhance the model's generalization capabilities across various datasets, particularly demonstrating significant advantages in datasets related to practical applications like fault detection. Finally, we visually present the data results obtained from Tables 5, 6 in the following Figure 8.

According to the data in Tables 7, 8, it is evident that our proposed model exhibits a significant advantage in computational resource metrics, including training time, inference time, and the number of model parameters across the four datasets. Specifically, on the ICS-SGAD dataset, our model's training time is 2.99 s less than Azad et al.'s model, which is the second-best after our own, and

TABLE 3 Comparison of Specificity, Accuracy, Recall and F1-score indicators in different methods based on the number of low energy discharge, high energy discharge and medium and low temperature overheating faults.

Model	Fault type											
	C1				C2				C3			
	Specificity (%)	Accuracy (%)	Recall (%)	F1-score	Specificity (%)	Accuracy (%)	Recall (%)	F1-score	Specificity (%)	Accuracy (%)	Recall (%)	F1-score
Belhadi, Asma et al	85.31	85.63	85.60	85.61	83.37	83.53	83.55	83.54	85.41	85.35	85.28	85.60
Andresen et al	86.55	86.51	86.94	86.72	85.84	83.16	84.73	83.94	86.18	86.08	86.91	86.49
He, Shunfan et al	87.14	87.39	87.97	87.68	85.93	84.19	84.76	84.47	88.48	86.77	87.36	87.06
Abrao, Taufik et al	87.97	87.96	88.44	88.20	86.15	85.62	85.97	85.79	89.07	87.48	88.00	87.74
Hosseinzadeh et al	88.56	88.58	89.65	89.11	88.64	86.95	87.54	87.24	89.81	87.58	89.19	88.38
Azad, Salahuddin et al	89.81	90.49	90.14	90.31	89.17	88.63	90.40	89.51	90.48	88.62	90.70	89.65
Ours	91.39	90.83	91.84	91.33	91.42	89.52	90.22	89.87	92.04	90.35	92.16	91.25

The bold values represent the best results.

TABLE 4 Comparison of Specificity, Accuracy, Recall and F1-score indicators in different methods based on high temperature overheating, partial discharge and fault-free type.

Model	Fault type											
	C4				C5				C6			
	Specificity (%)	Accuracy (%)	Recall (%)	F1-score	Specificity (%)	Accuracy (%)	Recall (%)	F1-score	Specificity (%)	Accuracy (%)	Recall (%)	F1-score
Belhadi, Asma et al	83.17	83.66	83.09	83.37	85.16	85.77	85.19	85.48	85.61	84.58	84.17	84.37
Andresen et al	85.94	83.59	84.42	84.00	86.46	86.31	86.56	86.43	86.11	86.25	86.66	86.45
He, Shunfan et al	85.72	84.04	84.67	84.35	86.73	87.23	87.73	87.48	87.94	88.18	87.80	87.99
Abrao, Taufik et al	88.94	85.13	85.55	85.34	87.30	88.20	87.94	88.07	88.68	88.98	88.97	88.97
Hosseinzadeh et al	89.05	87.68	87.18	87.43	89.07	89.30	88.79	89.04	89.94	89.68	89.18	89.43
Azad, Salahuddin et al	89.84	88.14	88.34	88.24	90.19	90.38	88.82	89.59	91.27	91.05	91.47	91.26
Ours	90.28	89.52	90.22	89.87	91.34	91.36	89.20	90.27	94.39	93.98	94.42	94.20

The bold values represent the best results.

TABLE 5 Comparison of Specificity, Accuracy, Recall and F1-score indicators in different methods based on ICS-SGAD and AKNN-SGFD data sets.

Model	Datasets							
	ICS-SGAD dataset				AKNN-SGFD dataset			
	Specificity (%)	Accuracy (%)	Recall (%)	F1-score	Specificity (%)	Accuracy (%)	Recall (%)	F1-score
Belhadi, Asma et al	83.14	83.17	83.59	83.38	84.08	84.63	84.23	84.43
Andresen et al	84.85	84.03	85.45	84.73	85.40	84.99	84.88	84.93
He, Shunfan et al	85.59	86.04	86.81	86.42	86.34	86.11	85.93	86.02
Abrao, Taufik et al	86.17	86.25	88.01	87.12	87.19	87.20	86.32	86.76
Hosseinzadeh et al	88.94	88.58	89.65	89.11	89.72	89.52	89.44	89.48
Azad, Salahuddin et al	89.48	89.08	90.01	89.54	91.42	90.36	91.37	90.86
Ours	90.39	90.83	91.84	91.33	92.01	91.55	92.92	92.23

The bold values represent the best results.

TABLE 6 Comparison of Specificity, Accuracy, Recall and F1-score indicators in different methods based on FFC-SG and OWA-SGFD data sets.

Model	Datasets							
	FFC-SG dataset				OWA-SGFD dataset			
	Specificity (%)	Accuracy (%)	Recall (%)	F1-score	Specificity (%)	Accuracy (%)	Recall (%)	F1-score
Belhadi, Asma et al	85.41	85.13	85.39	85.26	82.31	82.41	82.53	82.47
Andresen et al	86.78	86.72	86.70	86.71	82.96	82.72	83.96	83.34
He, Shunfan et al	87.18	87.14	87.11	87.12	83.42	83.42	84.88	84.14
Abrao, Taufik et al	88.20	87.69	88.40	88.04	84.39	84.02	85.93	84.96
Hosseinzadeh et al	89.94	89.09	90.27	89.68	86.48	86.15	85.95	86.05
Azad, Salahuddin et al	91.07	90.87	91.53	91.2	88.70	87.92	88.71	88.31
Ours	93.67	93.94	92.79	93.36	90.79	90.40	91.02	90.71

The bold values represent the best results.

5.57 s less than the third-ranking model by Hosseinzadeh et al. Our model also demonstrates a reduction of 6.05 m in inference time and a decrease of 47.96 million parameters. On the AKNN-SGFD dataset, our model's training time is 1.18 s less than Azad's model and 2.4 s less than Hosseinzadeh's model. Additionally, our model achieves a decrease of 8.91 m in inference time and a reduction

of 15.43 million parameters. For the FFC-SG and OWA-SGFD datasets, our model's metrics in training time, inference time, and the number of parameters are superior to other models. Notably, on the FFC-SG dataset, our model's training time is 10.68 s less than Andresen et al.'s model and 5.50 s less than Abrao et al.'s model, with an even more significant reduction of 15.17 m in inference time.

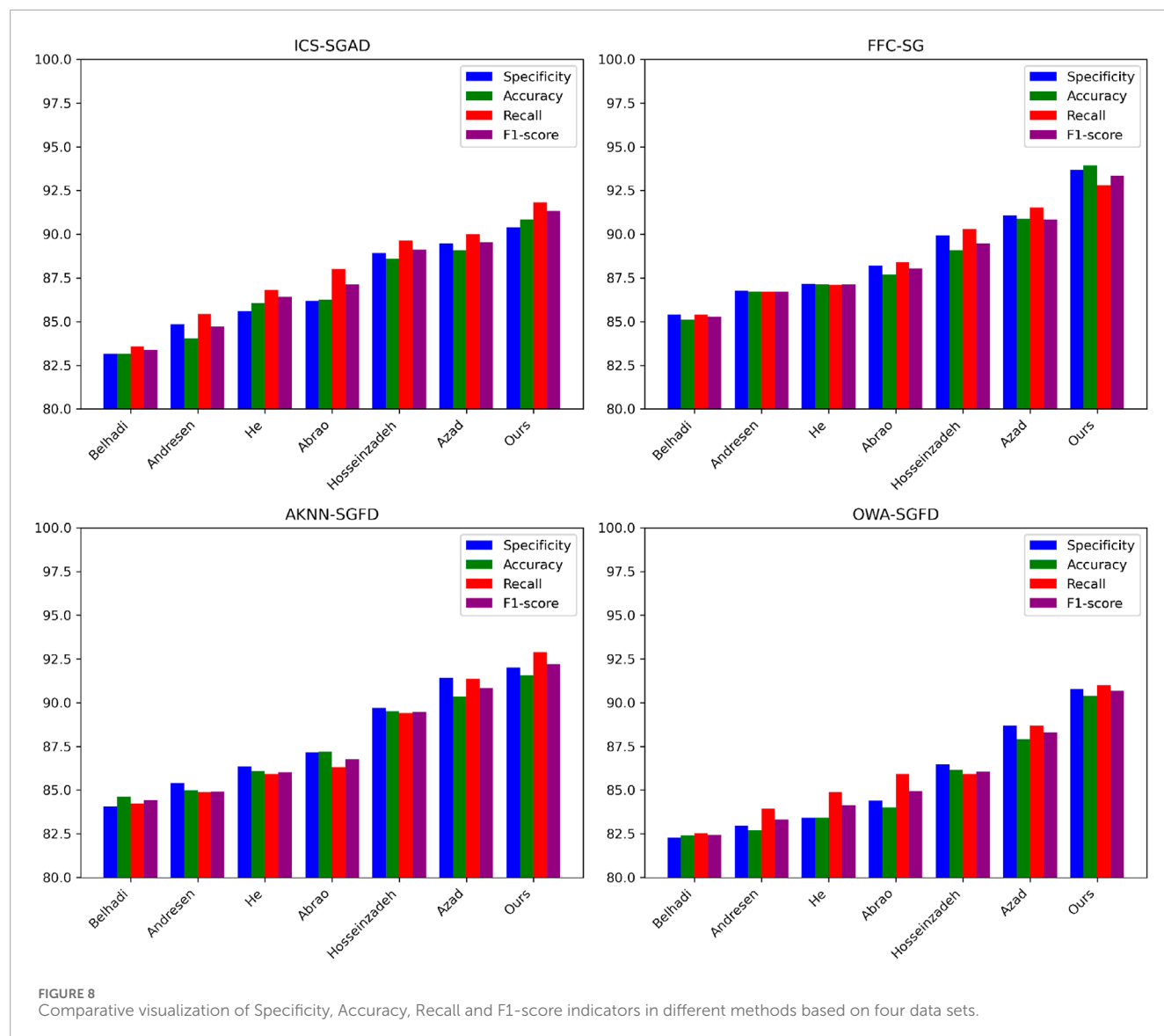


FIGURE 8 Comparative visualization of Specificity, Accuracy, Recall and F1-score indicators in different methods based on four data sets.

Overall, our model exhibits an average advantage of over 5% in each metric across the four datasets, highlighting the effectiveness of our approach in reducing computational resource consumption while maintaining or enhancing predictive performance across diverse datasets. This underscores the high practical value of our method in the application of equipment fault detection in electronic systems. Similarly, we visually represent the data results from Tables 7, 8 in the following Figure 9.

According to the data in Tables 9, 10, the effects of optimizing the model with different technical modules on the four datasets are evident. Compared to the baseline model, adding the transfer learning module can lead to some improvement in various metrics, but the enhancement is limited. For example, on the ICS-SGAD dataset, each metric, on average, only increases by approximately 13%. After incorporating the generative adversarial network module, the model performance experiences a more significant improvement, with an average increase of over 10% for each metric across the four datasets. However, our proposed approach,

combining transfer learning and GAN neural network modules, demonstrates the best results. This method leads to an average improvement of nearly 30% or more for each metric across the four datasets. Particularly noteworthy is the substantial improvement observed on the AKNN-SGFD and FFC-SG datasets, where each metric sees a significant increase. For instance, on the FFC-SG dataset, each metric improves by over 30%. This strongly indicates that the techniques we employed can comprehensively exploit dataset information by effectively combining transfer learning and adversarial learning approaches, significantly enhancing the model's generalization ability across various tasks. Compared to using transfer learning or GAN individually, our approach exhibits a clear advantage, providing strong evidence for its potential value in practical applications. Additionally, we visually present the data results from Tables 9, 10 in the following Figure 10.

According to the data in Tables 11, 12, the resource consumption of the model optimized with different technical modules on the four datasets is evident. Compared to the baseline model, adding

TABLE 7 Comparison of Training time, Inference time and Parameters indicators in different methods based on ICS-SGAD and AKNN-SGFD data sets.

Model	Datasets					
	ICS-SGAD dataset			AKNN-SGFD dataset		
	Training time(s)	Inference time(ms)	Parameters(M)	Training time(s)	Inference time (ms)	Parameters(M)
Belhadi, Asma et al	57.17	149.42	292.82	54.32	138.14	284.17
Andresen et al	54.47	142.47	287.45	52.15	134.72	268.37
He, Shunfan et al	51.51	137.16	275.25	49.93	131.84	250.50
Abrao, Taufik et al	48.40	130.74	267.52	46.17	127.37	246.41
Hosseinzadeh et al	47.75	124.68	260.74	43.29	122.27	241.94
Azad, Salahuddin et al	45.19	119.20	253.14	42.07	115.72	238.47
Ours	42.18	113.47	244.18	40.89	107.81	229.04

The bold values represent the best results.

TABLE 8 Comparison of Training time, Inference time and Parameters indicators in different methods based on FFC-SG and OWA-SGFD data sets.

Model	Datasets					
	FFC-SG dataset			OWA-SGFD dataset		
	Training time(s)	Inference time (ms)	Parameters(M)	Training time(s)	Inference time (ms)	Parameters(M)
Belhadi, Asma et al	52.92	129.96	275.74	58.17	151.27	297.18
Andresen et al	50.55	123.34	268.76	56.04	142.78	284.94
He, Shunfan et al	48.44	120.14	255.42	52.12	138.75	280.13
Abrao, Taufik et al	45.37	116.92	248.05	49.17	132.72	274.99
Hosseinzadeh et al	42.16	113.55	240.81	47.21	129.40	267.13
Azad, Salahuddin et al	41.12	108.71	234.39	44.47	124.13	259.90
Ours	39.87	101.75	227.41	43.91	112.02	248.38

The bold values represent the best results.

the transfer learning module can moderately reduce the model's training time, inference time, and parameter count. However, the reduction is limited. For example, on the ICS-SGAD dataset, the training time decreases by only 3.86 s, the inference time decreases from 147.75 m to 138.47 m, and the parameter count improves by 9.44 million. After incorporating the gan module, the model's computational costs decrease further, with training time and inference time both reducing by around 10% on all datasets, and a noticeable decrease in parameter count. The most effective results are achieved when we connect the transfer learning module and the generative adversarial network module in series. This approach leads to an average reduction of over 30% in training time, over 20% in inference time, and over 15% in parameter count across the four datasets. Particularly noteworthy is the more pronounced resource

savings on the FFC-SG and OWA-SGFD datasets. This indicates that our approach not only significantly improves model performance but also substantially reduces the model's computational costs, demonstrating its substantial practical value. Finally, we visually present the data results from Tables 11, 12 in the following Figure 11.

The model we proposed in this study outperforms the other six classical models on four different types of real datasets, both in terms of predictive performance indicators and computational resource consumption. Especially in practical application scenarios such as medical image recognition datasets FFC-SG and OWA-SGFD, our model's superiority is particularly evident, with an average improvement of over 5% or more in each metric and computational indicators compared to other models. This clearly demonstrates the advantages of our approach in these types of

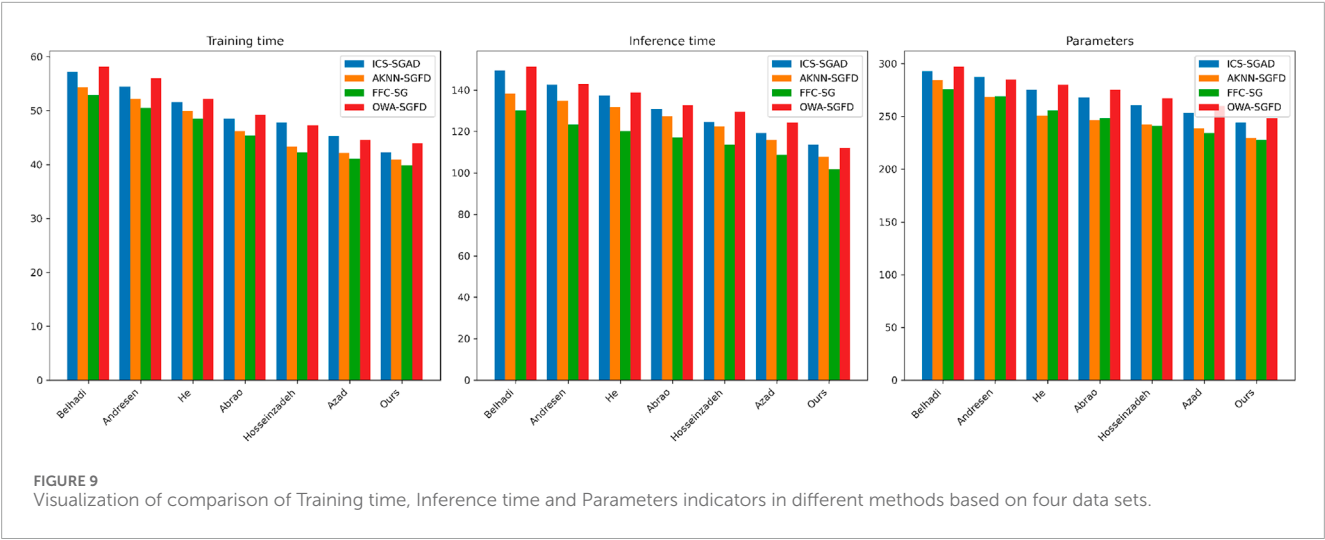


TABLE 9 Comparison of Specificity, Accuracy, Recall and F1-score indicators under different modules based on ICS-SGAD and AKNN-SGFD data sets.

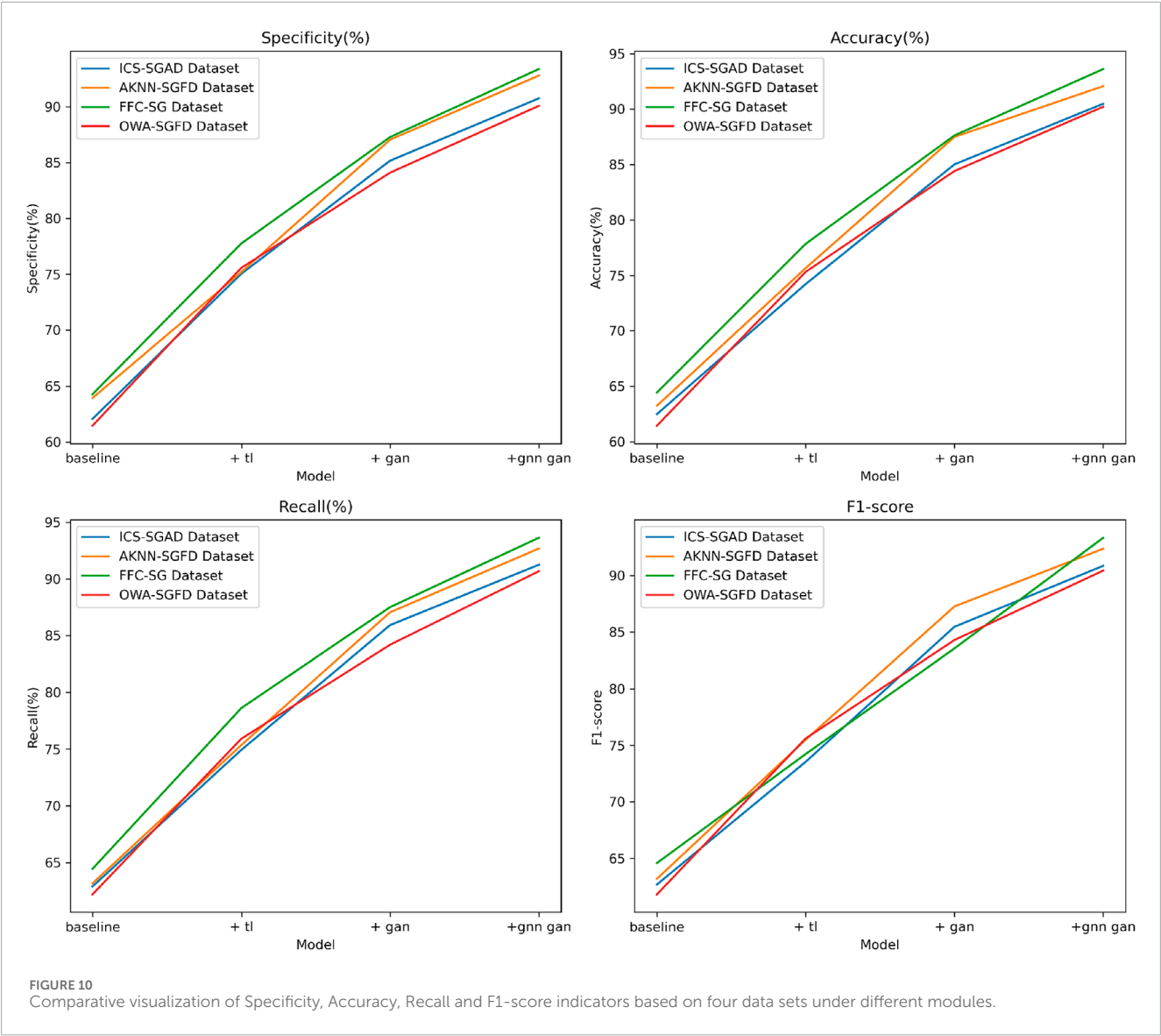
Model	Datasets							
	ICS-SGAD dataset				AKNN-SGFD dataset			
	Specificity (%)	Accuracy (%)	Recall (%)	F1-score	Specificity (%)	Accuracy (%)	Recall (%)	F1-score
baseline	62.07	62.48	62.90	62.69	63.94	63.24	63.17	63.20
+ tl	75.04	74.21	74.93	73.54	75.19	75.65	75.36	75.50
+ gan	85.16	85.01	85.93	85.47	87.04	87.49	87.06	87.27
+gnn gan	90.74	90.48	91.25	90.86	92.78	92.07	92.68	92.37

TABLE 10 Comparison of Specificity, Accuracy, Recall and F1-score indicators under different modules based on FFC-SG and OWA-SGFD data sets.

Model	Datasets							
	FFC-SG dataset				OWA-SGFD dataset			
	Specificity (%)	Accuracy (%)	Recall (%)	F1-score	Specificity (%)	Accuracy (%)	Recall (%)	F1-score
baseline	64.27	64.43	64.45	64.59	61.46	61.44	62.20	61.82
+ tl	77.75	77.84	78.62	74.23	75.57	75.31	75.92	75.61
+ gan	87.27	87.62	87.51	83.56	84.09	84.41	84.20	84.30
+gnn gan	93.36	93.61	93.63	93.34	90.07	90.21	90.68	90.44

tasks. Additionally, by comparing different technical modules, it can be observed that our combined approach of transfer learning and generative adversarial networks comprehensively exploits dataset information, significantly improving model performance, while also minimizing the computational burden associated with model training and inference. This holds significant value in industrial applications.

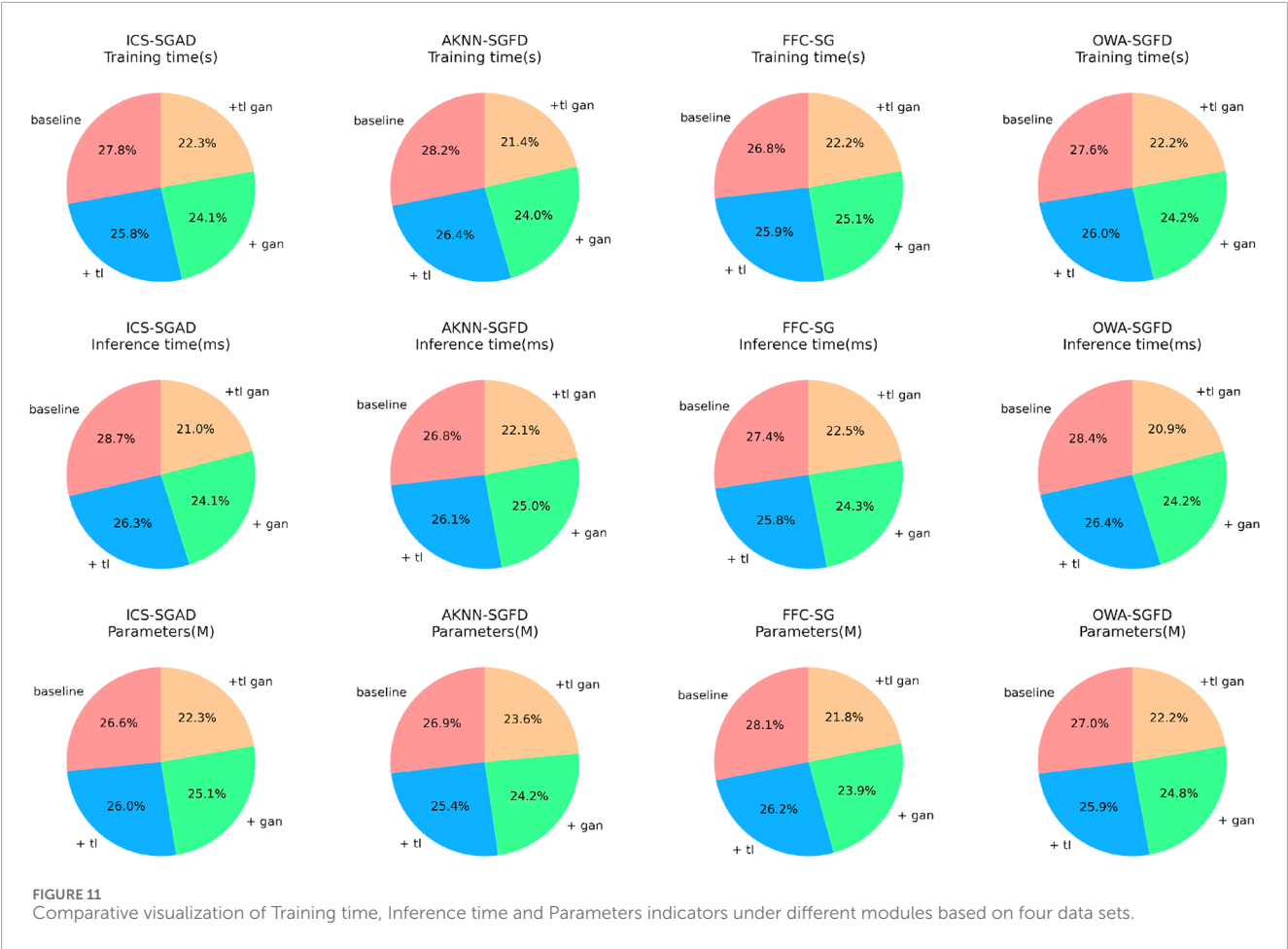
In summary, this series of experimental results thoroughly validates the wide applicability and outstanding performance of our proposed method in addressing real-world problems, providing valuable insights for relevant applications. We will continue to explore this direction in the future, aiming to continuously enhance the model's performance across various tasks and contribute to solving practical issues.



Model	Datasets					
	ICS-SGAD dataset			AKNN-SGFD dataset		
	Training time(s)	Inference time (ms)	Parameters(M)	Training time(s)	Inference time (ms)	Parameters(M)
baseline	54.29	147.75	266.15	50.42	138.27	250.97
+ tl	50.43	138.47	256.71	47.45	126.73	244.37
+ gan	47.06	126.18	248.58	44.12	116.37	234.04
+tl gan	43.48	112.45	220.27	40.48	101.08	206.51

TABLE 12 Comparison of Training time, Inference time and Parameters indicators under different modules based on FFC-SG and OWA-SGFD data sets.

Model	Datasets					
	FFC-SG dataset			OWA-SGFD dataset		
	Training time(s)	Inference time (ms)	Parameters(M)	Training time(s)	Inference time (ms)	Parameters(M)
baseline	50.17	141.42	251.08	51.33	138.88	269.94
+ tl	47.22	131.37	246.19	48.34	129.38	259.37
+ gan	44.60	120.34	237.39	46.05	118.02	248.09
+tl gan	41.27	104.08	210.47	44.90	107.67	222.40



4.6 Discussion

In this study, we employed a comprehensive approach by utilizing the Vision Transformer (ViT) model, transfer learning, and generative adversarial networks. Starting from the perspective of image processing, we seek to advance the digitization and intelligent perception levels of power systems. With the introduction of the ViT model, we accelerate the learning process through transfer learning and use generative adversarial networks to augment training data,

bringing significant technological breakthroughs to the field of power system image processing.

The key innovation of this research lies in the integration of multiple advanced technologies tailored to the practical requirements of emerging power systems. This integration aims to enhance the efficiency of UAV image recognition and intelligent power distribution network device fault detection. Through this study, we provide substantial support for the application of image processing and device fault detection in the construction of

emerging power systems, offering robust technical support for the development of the power industry, a focus of interest in related fields.

The significance of this study is manifested in several aspects. Firstly, by introducing the Vision Transformer (ViT) model, we have made significant progress in the field of power system image processing, providing a powerful tool for digital transformation. Secondly, the adoption of transfer learning effectively addresses the issue of limited data, enhancing the model's generalization ability and achieving more robust and reliable results in tasks such as power system equipment defect recognition. Lastly, the innovative application of generative adversarial networks in power system image recognition tasks contributes to augmenting training data, improving the model's accuracy in identifying various types of power equipment defects. This research is both theoretically innovative and strongly supported by experimental data, offering new insights for research and applications in the field of power system image processing.

Through experimental comparisons and analyses, our approach has achieved significant improvements in specificity, accuracy, recall, and F1-score metrics. Compared to traditional methods, our model excels in UAV image recognition and intelligent power distribution network device fault detection. Specifically, in the detection of faults in intelligent power distribution network devices, our model successfully captures subtle fault features, achieving a diagnostic accuracy of over 90%, an improvement of more than 17% compared to traditional methods. Additionally, the model demonstrates outstanding performance with an F1-score of around 91%. These experimental results not only validate the effectiveness of our approach but also showcase its enormous potential in practical Applications.

5 Conclusion

When discussing the continuous progress of technology, our research focuses on addressing challenges in the fields of unmanned aerial vehicle (UAV) image recognition and intelligent power distribution network device fault detection. This study aims to enhance recognition accuracy and efficiency by integrating cutting-edge technologies such as the Transformer model, transfer learning, and generative adversarial networks, providing innovative solutions for the safety and stable operation of power system equipment. In this chapter, we will review the research objectives, emphasize the innovation and importance of the study, outline the research findings, discuss limitations, and look ahead to future research directions. This research is dedicated to addressing practical issues, aiming to improve the performance of existing methods and offering new research perspectives for relevant areas in the field of emerging power systems.

Despite the considerable progress we've made in our research, there are still some limitations that require addressing. Firstly, our experimental dataset may not comprehensively cover all scenarios of power system images and distribution network device faults. As a result, the model's generalization performance could be limited in certain specific situations. Secondly, while we've employed generative adversarial networks to augment training data, further practical verification is necessary to evaluate the model's

generalization, particularly in extreme cases. These limitations underscore the importance of future research focusing on validation with more extensive datasets and enhancing the model's robustness and generalization capabilities. Additionally, the current Transformer model still faces challenges in terms of interpretability. To address this, we plan to introduce attention mechanisms in future research to enhance the model's interpretability and better explain its decision-making process.

Considering both the achievements and limitations of our study, future endeavors can focus on refining the model's performance and broadening its applicability. Firstly, we can delve into optimizing the model's hyperparameters to bolster its effectiveness in navigating complex scenarios. Secondly, validating our approach with more comprehensive datasets across diverse real-world application contexts will bolster its reliability. Furthermore, integrating advanced methodologies like self-supervised learning shows promise in elevating the model's performance, particularly in scenarios with scarce data.

In summary, our research successfully integrates cutting-edge technologies, including the Transformer model, transfer learning, and generative adversarial networks. This integration leads to significant advancements in unmanned aerial vehicle (UAV) image recognition and intelligent detection of distribution network device faults. Our approach not only holds promise in theory but also receives robust validation from experimental data. By pioneering the fusion of these technologies, we introduce important innovations in power system image processing. This lays a strong foundation for future research and applications in related domains. The outcomes of our study are poised to chart new paths in image processing and equipment fault detection within power systems, offering innovative solutions for the industry's digital transformation and intelligence. Through this endeavor, we establish a solid groundwork for further exploration and application in power system image processing. We are confident that these findings will catalyze innovative developments in related fields, providing invaluable technical support for the establishment and evolution of new power systems. We anticipate that our work will inspire fresh insights in the realm of novel power systems and serve as a conduit for collaboration and knowledge exchange between academia and the power industry.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

JZ: Conceptualization, Formal Analysis, Methodology, Resources, Writing—original draft, Writing—review and editing. YC: Investigation, Project administration, Supervision, Visualization, Writing—review and editing. JG: Methodology, Resources, Software, Validation, Writing—review and editing. XL: Investigation, Methodology, Project administration, Resources, Writing—original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. State Grid Chongqing Electric Power Co., LTD. Science and Technology Project: Research on Intelligent Fault Diagnosis Technology of Whole Chain of Intelligent iot System based on Artificial Intelligence (KJ2023_033).

Conflict of interest

Authors JZ, YC, JG, and XL were employed by Chongqing Electric Power Company, State Grid.

References

- Abid, A., Khan, M. T., and Iqbal, J. (2021). A review on fault detection and diagnosis techniques: basics and beyond. *Artif. Intell. Rev.* 54, 3639–3664. doi:10.1007/s10462-020-09934-2
- Appiah, A. Y., Zhang, X., Ayawli, B. B. K., and Kyeremeh, F. (2019). Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques. *Int. J. Photoenergy* 2019, 1–19. doi:10.1155/2019/6953530
- Ayoub, N., and Schneider-Kamp, P. (2021). Real-time on-board deep learning fault detection for autonomous uav inspections. *Electronics* 10, 1091. doi:10.3390/electronics10091091
- Azar, A. T., Koubaa, A., Ali, M. N., Ibrahim, H. A., Ibrahim, Z. F., Kazim, M., et al. (2021). Drone deep reinforcement learning: a review. *Electronics* 10, 999. doi:10.3390/electronics10090999
- Barta, G., Nagy, G. B. G., Kazi, S., and Henk, T. (2015). “Gefcom 2014—probabilistic electricity price forecasting,” in *Intelligent decision technologies: proceedings of the 7th KES international conference on intelligent decision technologies (KES-IDT 2015)*. Springer, 67–76.
- Chen, J., Chen, J., Zhang, D., Sun, Y., and Nanekaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. *Comput. Electron. Agric.* 173, 105393. doi:10.1016/j.compag.2020.105393
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: an overview. *IEEE signal Process. Mag.* 35, 53–65. doi:10.1109/msp.2017.2765202
- Fang, J., Li, H., Tang, Y., and Blaabjerg, F. (2018). On the inertia of future more-electronics power systems. *IEEE J. Emerg. Sel. Top. Power Electron.* 7, 2130–2146. doi:10.1109/jestpe.2018.2877766
- Giua, A., and Silva, M. (2018). Petri nets and automatic control: a historical perspective. *Annu. Rev. Control* 45, 223–239. doi:10.1016/j.arcontrol.2018.04.006
- Haji, S. H., and Abdulazeez, A. M. (2021). Comparison of optimization techniques based on gradient descent algorithm: a review. *PalArch's J. Archaeol. Egypt/Egyptology* 18, 2715–2743.
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., et al. (2022). A survey on vision transformer. *IEEE Trans. pattern analysis Mach. Intell.* 45, 87–110. doi:10.1109/tpami.2022.3152247
- Jalil, B., Leone, G. R., Martinelli, M., Moroni, D., Pascali, M. A., and Berton, A. (2019). Fault detection in power equipment via an unmanned aerial system using multi modal data. *Sensors* 19, 3014. doi:10.3390/s19133014
- Joshi, T., Goyal, N., and Ram, M. (2022). An approach to analyze reliability indices in peer-to-peer communication systems. *Cybern. Syst.* 53, 716–733. doi:10.1080/01969722.2022.2047273
- Karimulla, S., and Ravi, K. (2019). “A review on importance of smart grid in electrical power system” in *2019 international Conference on Computation of power, energy, Information and communication (ICCPIC)* (IEEE), 022–027.
- Kumar, N., Singh, B., and Panigrahi, B. K. (2023). Voltage sensorless based model predictive control with battery management system: for solar pv powered on-board ev charging. *IEEE Trans. Transp. Electrification* 9, 2583–2592. doi:10.1109/TTE.2022.3213253
- Kumari, P., Kumar, N., and Panigrahi, B. K. (2023). A framework of reduced sensor rooftop spv system using parabolic curve fitting mppt technology for household consumers. *IEEE Trans. Consumer Electron.* 69, 29–37. doi:10.1109/TCE.2022.3209974
- Lee, D., La, W. G., and Kim, H. (2018). “Drone detection and identification system using artificial intelligence,” in *2018 international Conference on Information and communication technology convergence (ICTC) (IEEE)*, 1131–1133.
- Li, Y., Yu, F., Cai, Q., Yuan, K., Wan, R., Li, X., et al. (2019). “Image fusion of fault detection in power system based on deep learning,” 22. *Cluster Computing*, 9435–9443. doi:10.1007/s10586-018-2264-2
- Lin, T., Wang, Y., Liu, X., and Qiu, X. (2022). A survey of transformers. *AI Open*.
- Mirjalili, S., and Mirjalili, S. (2019). “Genetic algorithm,” in *Evolutionary algorithms and neural networks: theory and applications*, 43–55.
- Mohammadi, F., Nazri, G. A., and Saif, M. (2019). “A fast fault detection and identification approach in power distribution systems,” in *2019 international conference on power generation systems and renewable energy technologies (PGSRET)*. IEEE, 1–4.
- Nagy, G. I., Barta, G., Kazi, S., Borbély, G., and Simon, G. (2016). Gefcom2014: probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. *Int. J. Forecast.* 32, 1087–1093. doi:10.1016/j.ijforecast.2015.11.013
- Neeraj, M. J., and Behera, R. K. (2022). Power load forecasting based on long short term memory-singular spectrum analysis. *Energy Syst.*, 1–23.
- Nguyen, H. T., Walker, C., and Walker, E. A. (2018). *A first course in fuzzy logic*. Chapman and Hall/CRC.
- Ning, X., Tian, W., He, F., Bai, X., Sun, L., and Li, W. (2023). Hyper-sausage coverage function neuron model and learning algorithm for image classification. *Pattern Recognit.* 136, 109216. doi:10.1016/j.patcog.2022.109216
- Ning, X., Yu, Z., Li, L., Li, W., and Tiwari, P. (2024). Dirl: differentiable rendering-based multi-view image-language fusion for zero-shot 3d shape understanding. *Inf. Fusion* 102, 102033. doi:10.1016/j.inffus.2023.102033
- Noda, T., Tadokoro, T., and Dozaki, T. (2023). Automatic generation of power system simulation data cases from utility databases: introducing a new technology. *IEEE Electrification Mag.* 11, 79–85. doi:10.1109/MELE.2023.3320521
- Peng, C., and Li, W. (2023). Adaptive spatial feature fusion object detection algorithm based on attention improvement. *J. Jilin Univ. Sci. Ed.* 61. doi:10.13413/j.cnki.jdxblxb.2022074
- Rahaman, M., Hillas, M. M., Tuba, J., Ruma, J. E., Ahmed, N., and Rahman, R. M. (2022). Effects of label noise on performance of remote sensing and deep learning-based water body segmentation models. *Cybern. Syst.* 53, 581–606. doi:10.1080/01969722.2021.1989171
- Ren, B., and Wang, Z. (2024). Strategic priorities, tasks, and pathways for advancing new productivity in the Chinese-style modernization. *J. Xian Univ. Finance Econ.* 37, 3–11. doi:10.19331/j.cnki.xjufe.20240008.002
- Scutari, M., and Denis, J. B. (2021). *Bayesian networks: with examples in R*. CRC Press.
- Simonyan, K., and Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.

Voita, E., Talbot, D., Moiseev, F., Sennrich, R., and Titov, I. (2019). Analyzing multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned. *arXiv Prepr. arXiv:1905.09418*.

Wang, J., Li, F., An, Y., Zhang, X., and Sun, H. (2024). Towards robust lidar-camera fusion in bev space via mutual deformable attention and temporal aggregation. *IEEE Trans. Circuits Syst. Video Technol.*, 1. doi:10.1109/TCSVT.2024.3366664

Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., et al. (2019). "Heterogeneous graph attention network," in *The world wide web conference*, 2022–2032.

Yao, Y., and Liu, Z. (2024). The new development concept helps accelerate the formation of new quality productivity: theoretical logic and implementation paths. *J. Xi'an Univ. Finance Econ.* 37, 3–14.

Zhao, S., Blaabjerg, F., and Wang, H. (2020). An overview of artificial intelligence applications for power electronics. *IEEE Trans. Power Electron.* 36, 4633–4658. doi:10.1109/tpel.2020.3024914

Zhu, H., Li, H., Li, W., and Li, F. (2021). Single image super-resolution reconstruction based on generative adversarial network. *J. Jilin Univ.* 59, 1491–1498.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A comprehensive survey on transfer learning. *Proc. IEEE* 109, 43–76. doi:10.1109/jproc.2020.3004555

Frontiers in Energy Research

Advances and innovation in sustainable, reliable
and affordable energy

Explores sustainable and environmental
developments in energy. It focuses on
technological advances supporting Sustainable
Development Goal 7: access to affordable,
reliable, sustainable and modern energy for all.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact



Frontiers in Energy Research

