frontiers | Research Topics

Data-driven approaches for efficient smart grid systems

Edited by
Jinran Wu, Yang Yang, Shaolong Sun and Yang Yu

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 978-2-8325-5909-3
DOI 10.3389/978-2-8325-5909-3

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

Data-driven approaches for efficient smart grid systems

Topic editors

Jinran Wu — The University of Queensland, Australia
Yang Yang — Nanjing University of Posts and Telecommunications, China
Shaolong Sun — Xi'an Jiaotong University, China
Yang Yu — Nanjing University of Posts and Telecommunications, China

Citation

Wu, J., Yang, Y., Sun, S., Yu, Y., eds. (2025). *Data-driven approaches for efficient smart grid systems*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5909-3

Table of contents

frontiers | Frontiers in Energy Research

TYPE Editorial
PUBLISHED 06 January 2025
DOI 10.3389/fenrg.2024.1536459

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
ZhaoYang Dong,
City University of Hong Kong, Hong Kong SAR, China

*CORRESPONDENCE
Jinran Wu,
ryan.wu@acu.edu.au

RECEIVED 28 November 2024
ACCEPTED 13 December 2024
PUBLISHED 06 January 2025

CITATION
Wu J, Yang Y, Sun S and Yu Y (2025) Editorial: Data-driven approaches for efficient smart grid systems.
Front. Energy Res. 12:1536459.
doi: 10.3389/fenrg.2024.1536459

COPYRIGHT
© 2025 Wu, Yang, Sun and Yu. 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.

Editorial: Data-driven approaches for efficient smart grid systems

Jinran Wu[1]*, Yang Yang[2], Shaolong Sun[3] and Yang Yu[2]

[1]Australian Catholic University, Banyo, Australia, [2]Nanjing University of Posts and Telecommunications, Nanjing, China, [3]Xi'an Jiaotong University, Xi'an, China

KEYWORDS

statistical modeling, deep learning, computational intelligence, metaheuristics algorithm, optimizations

Editorial on the Research Topic
Data-driven approaches for efficient smart grid systems

Smart grid systems (SGSs) are leading the modernization of energy infrastructure by integrating advanced technologies to improve efficiency, reliability, and sustainability. These systems demand sophisticated tools to address their complexity, with forecasting and optimization being crucial areas of focus. Machine learning (ML) techniques, including both traditional neural networks and advanced deep learning approaches, play a significant role in tackling the intricate challenges of SGSs. These methods enable accurate forecasting, which is essential for predicting electricity demand, renewable energy generation, and system loads. By supporting informed decision-making and efficient resource allocation, ML provides both theoretical contributions and practical applications for smart grids. This Research Topic, *Data-Driven Approaches for Efficient Smart Grid Systems*, explores the innovative use of machine learning to address challenges specific to SGSs. Forecasting is central to these efforts, as it forms the basis for understanding and managing the complex dynamics of SGSs. While traditional methods have demonstrated promise, they also highlight limitations in adaptability, scalability, and precision, particularly when addressing the evolving needs of modern smart grids. These challenges call for advanced algorithms that integrate diverse data sources, capture spatiotemporal relationships, and account for uncertainties.

The Research Topic is organized into four thematic areas ("forecasting and prediction techniques", "optimization and scheduling in power systems", "data quality, validation, and identification", and "research trends and evaluations in energy systems"), which highlight the variety of approaches and contributions from the 13 papers accepted.

The first area focuses on forecasting and prediction techniques, essential for managing renewable energy sources and optimizing grid operations. Yang et al. presented "*A Study of Short-Term Wind Power Segmentation Forecasting Method Considering Weather on Ramp Segments*," introducing a hybrid method that combines LightGBM-LSTM for non-ramping segments and CNN-BiGRU-KDE for ramp segments, achieving improved accuracy under extreme weather conditions. Zhao et al. proposed "*Photovoltaic Output Prediction Based on VMD Disturbance Feature Extraction and WaveNet*," which integrates variational mode decomposition with WaveNet, achieving a mean absolute percentage error of 6.94%. Huang et al. introduced "*Ultra-Short-Term Prediction of Microgrid Source Load*

Power Considering Weather Characteristics and Multivariate Correlation," presenting a joint forecasting model that uses multivariate correlation techniques and neural networks to enhance prediction accuracy for microgrid sources and loads. Peng et al. proposed "*Short-Term Wind Power Prediction and Uncertainty Analysis Based on VDM-TCN and EM-GMM*," using a VDM-TCN-based method to reduce RMSE errors and Gaussian mixture models for uncertainty analysis. Chen et al., in "*A Physical Virtual Multi-Graph Convolutional Coordinated Prediction Method for Spatio-Temporal Electricity Loads Integrating Multi-Dimensional Information*," developed a multi-graph convolutional network that models spatiotemporal dependencies, leading to more accurate short-term electricity load forecasts.

The second area explores optimization and scheduling in power systems, which are critical for ensuring cost-effective and reliable operations. Lu et al., in "*Optimal Scheduling of the Active Distribution Network with Microgrids Considering Multi-Timescale Source-Load Forecasting*," presented a hierarchical scheduling strategy that combines CNN-LSTM forecasting with a two-layer optimization framework, reducing operational costs and network losses. Jain and Gupta contributed "*Optimal Placement of Distributed Generation in Power Distribution Systems and Evaluating the Losses and Voltage Using Machine Learning Algorithms*," introducing a hybrid ML approach that combines support vector machines, random forest, and radial neural networks to optimize distributed generation placement, improving voltage profiles and minimizing power losses.

The third area addresses data quality, validation, and identification, which are fundamental for the success of ML applications in SGSs. Jiang et al., in "*A Multi-Task Learning-Based Line Parameter Identification Method for Medium-Voltage Distribution Networks*," proposed a framework that uses graph attention networks and multi-gate mixture-of-experts to improve line parameter identification. Hu et al. developed "*Bad Data Identification Method Considering the On-Load Tap Changer Model*," a two-stage method that addresses nonlinearities introduced by on-load tap changers, improving state estimation reliability in distribution networks. Cao et al., in "*Intelligent Substation Virtual Circuit Verification Method Combining Knowledge Graph and Deep Learning*," proposed an approach combining knowledge graphs with deep learning to improve the accuracy of virtual circuit validation in substations. Li et al., in "*Channel Prediction Method Based on Data-Driven Techniques for Distribution Automation Main Stations*," introduced an adaptive broad learning network for communication channel state prediction, enhancing transmission quality.

The final area examines research trends and evaluations in energy systems, focusing on broader developments and identifying gaps in energy system research. Jain and Gupta, in "*Evaluation of Electrical Load Demand Forecasting Employing Various Machine Learning Algorithms*," evaluated short-term power load prediction using 5 years of data from the Chandigarh UT electricity utility. The study compares algorithms such as LSTM, SVM, ensemble classifiers, and deep learning methods, concluding that LSTM achieves the lowest prediction errors, outperforming SVM by 13.51%. Kizielewicz, in "*Onshore Power Supply: Trends in Research Studies*," examined the development of Onshore Power Supply (OPS) systems in ports to reduce exhaust gas emissions. By analyzing technical, economic, and logistic factors, the study identifies key research gaps and outlines a roadmap for future exploration.

The contributions in this Research Topic demonstrate the potential of machine learning to advance smart grid systems. By addressing challenges in forecasting, optimization, data validation, and broader research trends, these papers provide valuable insights and practical solutions for improving the efficiency, reliability, and sustainability of SGSs. This collection reflects the growing importance of data-driven approaches in bridging the gap between research and real-world applications. We hope these works inspire further research and collaboration across academia, industry, and policymaking to enhance the capabilities of smart grids and support global energy sustainability.

Author contributions

JW: Writing–original draft, Writing–review and editing. YYa: Writing–review and editing. SS: Writing–review and editing. YYu: 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.

frontiers | Frontiers in Energy Research

TYPE Review
PUBLISHED 15 March 2024
DOI 10.3389/fenrg.2024.1383142

Check for updates

OPEN ACCESS

EDITED BY
Jinran Wu,
Australian Catholic University, Australia

REVIEWED BY
Chanjuan Liu,
Shanghai Customs College, China
Nishant Kumar,
Indian Institute of Technology Jodhpur, India

*CORRESPONDENCE
Joanna Kizielewicz,
j.kizielewicz@wznj.umg.edu.pl

RECEIVED 06 February 2024
ACCEPTED 05 March 2024
PUBLISHED 15 March 2024

CITATION
Kizielewicz J (2024), Onshore power supply–trends in research studies.
Front. Energy Res. 12:1383142.
doi: 10.3389/fenrg.2024.1383142

COPYRIGHT
© 2024 Kizielewicz. 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.

Onshore power supply–trends in research studies

Joanna Kizielewicz*

Department of Management and Economics, Faculty of Management and Quality Science, Gdynia Maritime University, Gdynia, Poland

Restrictive regulations introduced by the European Parliament and the United Nations have forced the seaport authorities to reach net zero gas emissions by 2030. An important source of pollution emitted in ports involves exhaust gas emissions from ships powered by ship engines while they are berthed in seaports. One of the ways to reduce the level of pollution and nuisance caused by ship engines is to equip ships with Onshore Power Supply (OPS). Unfortunately, still few ports can boast this type of systems at their quays. There are many publications devoted to the analysis of this phenomenon in the technical and engineering terms, but a significant part also concerns the economic, financial, logistic and organizational aspects. The aim of this research is to identify the directions of development as for research into OPS, to organize the terminology devoted to OPS, and to define scientific disciplines of research in the field of OPS. This research was conducted on the basis of the data mining method using the Scopus, EBSCO and Web of Science databases, applying the multi-stage selection of criteria. The research results may provide an interesting material for scientists to identify the gaps of research regarding OPS.

KEYWORDS

port management, onshore power supply, sustainable development, smart grid, energy efficiency

1 Introduction

The seaport authorities are currently facing a major challenge related to investments supporting the protection of environment and counteracting its pollution. Changes in the Earth's climate resulting from heavy environmental pollution, caused to a large extent by the activities in the maritime economy, have made international institutions and organizations raise the alarm and introduce numerous restrictions regulated by law. Next to the International Maritime Organization (IMO) Convention for the Prevention of Pollution from Ships (MARPOL) (IMO, 1973), It is worth mentioning primarily the most important ones, i.e., The Directive 2000/59/EC on port reception facilities for ship-generated waste and cargo residues with subsequent additions; Annex VI MARPOL Convention regarding limits on emissions of SOx, NOx and particulate matter from exhaust gases emitted from ships (IMO, 2015), and also a document called "Transforming our world: the 2030 Agenda for Sustainable Development" (General Assembly, 2015). In addition, two other documents developed in recent years, i.e.,: 17 Sustainable Development Goals (SDGS) issued in 2015 by the Organization of United Nations, and the European Green Deal (EC, 2019) issued in 2019 by the European Commission should also be mentioned.

The above-mentioned documents specify, *inter alia*, the permissible limits for pollutant emissions in maritime transport and when ships are berthed in ports, and pave the way for zero. In the case of seaports, it primarily refers to creating conditions aimed to reduce the

level of pollution emitted by ships in ports (Viana et al., 2014) increase the energy efficiency in port facilities and provide it for the operators of ships at berth as well. Seaports are taking various measures in this regard, and among them there is one solution that is lately gaining in popularity, i.e., Onshore Power Supply (OPS). (Vaishnav et al., 2016; Winkel et al., 2016; Innes and Monios, 2018; Chen et al., 2019; OPS Master Plan, 2021; World Port Sustainability Program, 2021). Recently, the area of research as regards OPS has been very popular, but it should be noted that the source literature provides various definitions for OPS, and they are discussed below in the first part of the study.

The above-mentioned issues and many other related to OPS are currently the subject of numerous studies and analyses of scientists from around the world. Therefore, the aim of this paper is to identify the directions of OPS research development and to define the scientific disciplines of research in the field of OPS. The following research questions were formulated: 1) What are the benefits of investments in OPS? 2) In which scientific disciplines is OPS research currently conducted? 3) What are the current trends and areas of research devoted to OPS?

The content of this paper is typical of such scientific papers; it means that the paper consists of five main sections, including an introduction to the subject related to the issues concerning OPS. The second part of this article presents the review of current achievements in the area of studies regarding OPS. The methodology of the systematic literature review applied in the analysis was described in the third part of this article, and the fourth part was devoted to the presentation of the results. The article ends with a discussion and conclusions expressing the recommendations for further research for the potential groups of interests as well as a list of reference literature.

In the article the method of a multi-stage systematic literature review with the use of data mining was applied. The process of selecting the criteria for analysis and the stages of research process are presented in the chapter devoted to methodology. This article may constitute the subject of interest for seaport authorities and cruise ship owners.

2 Literature review

The International Association of Ports and Harbours (IAPH) World Port Climate Initiative (WPCI) has developed an important document on Onshore Power Supply, which is a key source of information for all entities interested in investing in and using OPS, showing the benefits of replacing onboard-generated power from diesel auxiliary engines with electricity generated onshore (IAPH, 2010). In addition, the international standards for High-Voltage Shore Connection (HSCV) (ISO/IEC/IEEE 80005-1:2012, 2012) have been developed, describing systems for suppling ships with electrical power from shore.

The website provides information on the environmental benefits of OPS and the associated costs. It highlights the existing installations in ports worldwide and the existing suppliers of the OPS technology and overall provides guidance for OPS implementation. The website is primarily targeted to port authorities, terminal operators and shipping companies who are considering the introduction or expansion of OPS technology.

According to the International Maritime Organization (IMO) Shore Power Supply (OPS) "is considered a measure to improve air quality in ports and port cities, to reduce emissions of air pollutants and noise and, to a lesser extent, to reduce carbon dioxide through ships at berth replacing onboard generated power from diesel auxiliary engines with electricity supplied by the shore" (International Maritime Organization IMO, 2012). At this point, however, it should be noted that in the literature there are several terms defining OPS, i.e.,

- Onshore Power Supply (OPS) (Li and Du, 2020),
- Shore-To-Ship Power (SSP) (Innes and Monios, 2018),
- Cold Ironing (CI) (Ballini and Bozzo, 2015),
- Shore Side Electricity (SSE) (Winkel et al., 2016; Stolz et al., 2021),
- High-Voltage Shore Connection (HSCV) (ISO/IEC/IEEE 80005-1:2012, 2012), and
- Alternative Maritime Power (AMP) (Chen et al., 2019).

Bouman et al. (2017) indicated that Shore Side Electricity (SSE) is an effective CO^2 reduction measure for ships. Additionally, (Stolz et al., 2021), proved that "shore side electricity can drastically reduce the emissions from fossil fuel-powered auxiliary engines of ships at berth". They published very interesting research results presenting the "CO^2 emissions from the production of electricity required for SSE and compared to the emissions of ships at berth". (Stolz et al., 2021).

Cold-Ironing (CI) is another frequently used term for OPS. For example (Ballini and Bozzo, 2015), wrote that CI is "a fully developed technology that allows vessels at berth to use shore power rather than rely on electricity generated by their Auxiliary Engines (AE)". Moreover (Bakar et al., 2022), defined cold ironing as "an electrification alternative in the maritime sector used to reduce shipborne emissions by switching from fuel to electricity when a ship docks at a port". In this paper, the authors presented the method of forecasting ship berthing duration that can contribute to track of the cold ironing consumption. The forecasting method can be used by other seaports in estimating demand.

These various terms used to describe OPS have been applied as keywords in the data mining process among scientific databases.

Various technological solutions for OPS solutions for ships are currently offered on the market, i.e.,: mobile power generator units, compact modular cabling system, or main transformer station with local stations at the berths, and also shore power box and other. It is worth emphasizing that the costs of OPS installations depend on (Bullock et al., 2023):

- types of OPS unit technology (e.g., centralised unit, outlet points, transformer, frequency converter, groundworks etc.);
- shore network ancillary equipment;
- distance from the power source of OPS devices;
- power draw required for ships and OPS devices;
- cable management system;
- types of vessels moored at the berths in seaports;
- berthing capacity, and
- many other factors.

The average cost of the OPS installation varies between €1 million and €35 million, e.g., a total budget of 36.4 million US dollars was invested in the OPS installation in the Port of Bremen/Bremerhaven (The Port of Bremen, 2023). Moreover, important are also the costs of electricity transportation to the OPS devices, which depend on.

- port location;
- accessibility to the berths from land and sea;
- sources of energy generation;
- power demand;
- vessels types;
- frequency of cruise ships calls.

Low-power ferries and, for example, small cruise ships with the capacity up to 1000 passengers (Pax) require power only < 3 MW (megawatt), but giant cruise ships with the capacity of 10,000 Pax onboard require power from 6 MVA (megawoltamper) to 20 MVA. The average cost of energy transportation can amount from $300,000 to $4 million per berth. (Bullock et al., 2023; The Port of Bremen, 2023) [6969].

It is also crucial to maintain OPS, and the related costs depend on many factors, i.e.,: costs of maintenance of OPS infrastructure, price of electricity, number of ships mooring into berths (calls), and time of mooring ships into berths (hours).

For more than a decade, scientific research has been conducted in various fields of OPS science by scientists representing technical, engineering, geographical, biological, chemical sciences, as well as social sciences in the field of economics and management (Puig et al., 2022). Maritime transport causes huge environmental pollution. As research shows among global atmospheric pollutants, 2 percent of carbon dioxide (CO_2), 10 to 15 percent of nitrogen oxides (NO_x) and 6 percent of sulfur oxides (SO_x) come from ships. Solutions that will reduce the scale of this phenomenon have been sought (Stolz et al., 2021). OPS is just one of such solution that can bring numerous benefits to the region, local communities, as well as ship owners. The use of OPS in ports provides, among others: lower fuel consumption by ships, reduction of pollution emitted by ships, ship engine noise reduction, removal of vibration caused by ship engines, and easier maintenance of water purity.

Moreover, disconnecting the ships from ship engine power supply and connecting them to the onshore generator also involves savings for the shipowners. However, one more issue needs to be raised, namely, the fact that still few vessels are equipped with devices for connecting the ships to onshore power supply. Despite obvious environmental benefits resulting from using OPS in seaports, many port authorities are still reviewing the economic effectiveness of these solutions. Investments in OPS are highly capital-intensive and their return is spread out over time. The first economic reports indicate that the investments in OPS in ports do not always show satisfactory economic efficiency, but they are certainly ecologically responsible and socially justified (Ballini and Bozzo, 2015).

A review of previous achievements in the field of research on energy system models and onshore power supply systems has showed that this area of research is studied by researchers around the world, i.e.,: in China (Chen et al., 2019), United States, (Bouman et al., 2017; Bullock et al., 2023), Norway (McArthur and Osland 2013), Germany (World Port Sustainability Program, 2021), Spain (OPS Master Plan, 2021). and other European seaports (Winkel et al., 2016). For example, the team (Mattsson et al., 2021) studied, among others: the factors for onshore and offshore wind power, and existing and future hydropower as part of research on the Energy System Model. The authors noted that in order to develop an effective energy system model, cross-sectoral cooperation is necessary and research in this area should also be carried out.

In 2016, (Winkel et al., 2016). raised in their research the economic and environmental aspects of the use of OPS, but they used the term Shore Side Electricity (SSE). Innes and Monios (2018) presented analyses regarding the data on energy demand, and presents a few scenarios for small and medium size ports in the field of cold ironing installation. Qi et al. (2020) showed the results of analysis concerning shore power economic challenges from the perspective of different groups of interests including, i.e., ship owners, port authorities, and state.

Interesting research on policy support on both capital funding and tax reform, and also factors affecting shore power economics for ship and port operators presented by the team of Bullock et al. (2023) based on research in port of Aberdeen in Scotland. The authors raised an important issue with OPS, i.e., "high capital costs for ports, high taxes on land-side electricity and the global lack of taxation on ships' fuel oils" (Bullock et al., 2023) and pointed to a gap in research on this issue. This work is particularly worthy of deeper analysis, because it contains both economic aspects and feasibility study, including, *inter alia*, potential shore power demand by ship type and economic viability, and what is crucial environmental and social valuation. Bouman et al. (2017) presented a comprehensive review of CO_2 studies, emissions, reductions, potentials and measures. Many works address the issues of shipping decarbonisation. OPS installations are one of the solutions that is just analyzed by the author.

An important element of research that should also be taken into account is the development of energy facilities in the vicinity of seaports enabling the connection to OPS. Interesting studies on offshore energy hubs were presented by Zhang et al. (2022). The authors presented the model that "can be used to analyse the interaction of an offshore energy system and onshore energy system transition" (Zhang et al., 2022) It should be emphasized that OPS should be powered by renewable energy sources. It is crucial in the decarbonization process.

Current research studies regard the use of new technologies to optimize processes in electrification. The really interesting research in this area has been carried out, among others, by the research team (Wang et al., 2021b) who presented the design principles and techniques used to implement ternary logic gates using memristor-CMOS hybrid chips for use in strategies for optimizing performance indicators, including, e.g., power consumption. A similar line of research was also carried out by Su et al. (2023) presents innovative research presenting ways to monitor the condition of batteries, based on the integration of physical modelling with machine learning techniques. These types of solutions allow for monitoring and evaluation of battery performance and can be used in optimising the performance of batteries and in projects related to the electrification of port devices.

Another interesting field of research concern the safety assurance was carried out by Siu et al. (2022) who presented the use of a multi-agent command authentication system as a solution to increase security and economic reliability in power systems, with particular emphasis on the possibility of emerging cyberattacks.

The important area of research studies in the field of OPS relates process of integration of energy systems with various power sources. An interesting study in this area was carried out by the team of researchers V. Saxena et. al. (2021) (Saxena et al., 2012) who proposed the MPC Based Algorithm and its validation method for a multifunctional grid-integrated photovoltaic system. The results demonstrate the effectiveness of the model-based predictive control (MPC) algorithm for ensuring energy quality and reliability in the integration of renewables into the power grid. Similar studies relating to the management of the integration of renewable energy sources and the operation of microgrids were carried out by the team of N. Kumar et. al. (Kumar et al., 2018). The paper proposes an intelligent control strategy with the use of artificial intelligence to effectively synchronize the microgrid with the main power grid, taking into account the situation of partial shading. Thanks to the proposed solutions, it is possible to ensure energy efficiency and thus reduce economic and environmental costs. On the other hand, research supporting practical solutions for the control and limitations of control techniques for complex systems in various network structures and dynamic conditions was presented by G. Chen (2023) (Chen, 2022). To sum up, it can be seen that onshore power supply is the subject of numerous scientific studies of various nature and range, both geographical and thematic and it seems reasonable to try further in-depth analysis.

3 Methodology of the systematic literature review

The research was a type of qualitative research. In the article, the method of a multi-stage systematic literature review with the use of data mining was applied. The independent variable is Onshore Power Supply (OPS). The literature analysis process consisted of six stages. To achieve the intended objective of the analysis, i.e., to identify the development of OPS research and to organize the terminology devoted to OPS, in the first stage, key phrases and key words related to onshore power supply were defined. In the second stage, three databases were selected to conduct the search, i.e., Science Direct, EBSCO and Web of Science. In the third stage of analysis, selection criteria that narrowed the search to specific areas were defined. The first criterion involved defining the period of search time, which was limited to a decade, i.e., from 2014 to 2023. In addition, the type of publications to be analysed were specified; only research papers and review papers available in the Open Access option were included. Using publications with limited access to conduct the extensive data analysis, as is the case here, is too expensive and often exceeds the researchers' budgets.

In the next stage of analysis (the fourth), the search was narrowed down to the phrases listed in No_x table in the titles of publications, keywords and abstracts from these publications. As a result, a list of publications for detailed analysis was defined. In the fifth stage, further criteria narrowing the search area were indicated and the collected database was grouped into four subject areas, i.e.,: 1) Engineering analysis (technology, technical, etc.) 2) Socio-economic analysis (financial, fiscal, business, management, etc.), 3) Environmental analysis (protection, pollution, emissions, etc.), 4) Legal analysis (legal standards, law regulations, etc.) (Scheme 1).

The full texts of selected publications were verified, selected and put through detailed substantive analysis, and the results thereof are presented in the next part of this study.

To analyse the data mining of the databases results, the critical and comparative analysis was applied, as well as inductive and deductive reasoning.

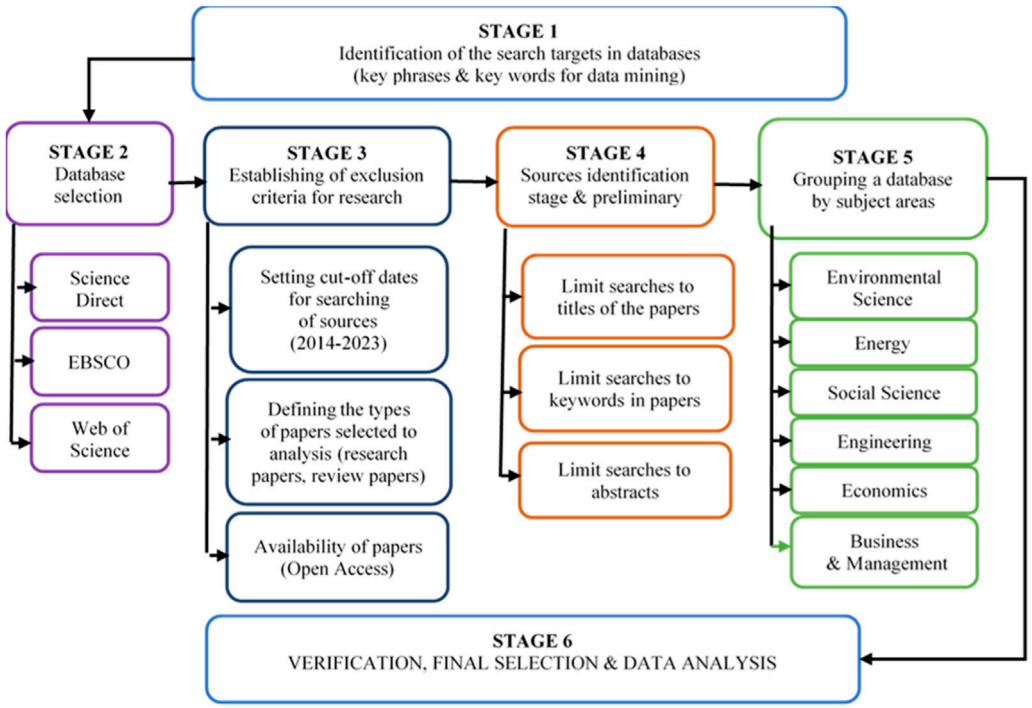
4 Discussion

A systematic review of databases showed that between 2014 and 2023, assuming the previously described selection criteria, a total of 74,111 scientific publications related to OPS were registered. The EBSCO database included a total of n = 70,424 scientific publications placed in review papers and research papers, which referred directly or indirectly to the issue of OPS, which accounted for as much as 94.98% of all publications in the examined scientific databases. Meanwhile, ScienceDirect database registered a total of 3431 publications (i.e., 4.63% of all publications) during the period under analysis. Unfortunately, in the Web of Science database there are few scientific publications related to OPS, i.e., 279, which accounts only for 0.39% of the total number (Figure 1).

A review of the ScienceDirect, EBSCO and Web of Science scientific databases also indicated that the majority of publications related to OPS, i.e., 63.47% is devoted to issues under the heading of "STSP" and 32.38% of "OPS". This is valuable information for scientists who will take up the subject of OPS in their future scientific work. The search limited in the literature queries only to the common (as it might seem) terms, i.e.: "OPS" (Onshore Power System) and neglecting "STSP" (Shore To-Ship Power) could significantly distort the scope of analysis. Other terms related to OPS, i.e., CI, SSE, or HVSC are less commonly used in the scientific literature (Figure 2).

By comparing the resources of the three scientific databases above-mentioned, in the area of publications devoted to OPS with the use of criteria described above, we also come to the conclusion that the largest number of publications devoted to "Shore To-Ship Power" and "Onshore Power Supply" was registered in the EBSCO database (94.98%). Similarly, in the ScienceDirect database, the largest number of publications were devoted to "Onshore Power Supply" and "Shore To-Ship Power". Whereas scientists rather rarely use other terms, i.e.: "Cold ironing", or "Alternative Maritime Transport" (Figure 3).

As a result of the analysis, a total of 74,135 publications were filtered from all three examined databases where references to OPS were found (3rd Stage). Then, in 4th Stage, in accordance with the planned model of systematic literature review, the search was narrowed only to publications referring directly to specific phrases in the titles of these publications and abstracts and keywords. This resulted in a reduction in the number of publications to 302. The next step in the 5th Stage was to group the publications into individual subject areas. In the last stage of the analysis (6th Stage) the publications were verified and selected. After



SCHEME 1
Database mining and analysis process. Source: own elaboration.

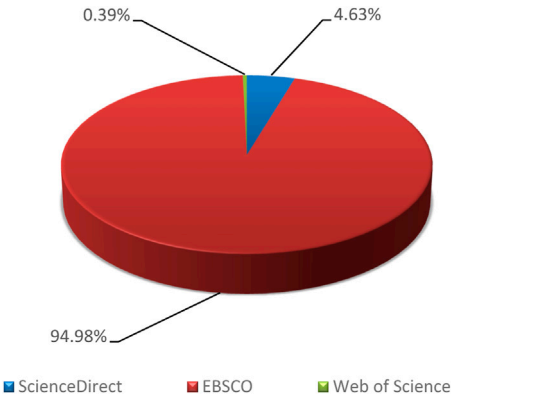


FIGURE 1
Percentage structure of the examined scientific databases by publications devoted to Onshore Power System. Source: own study.

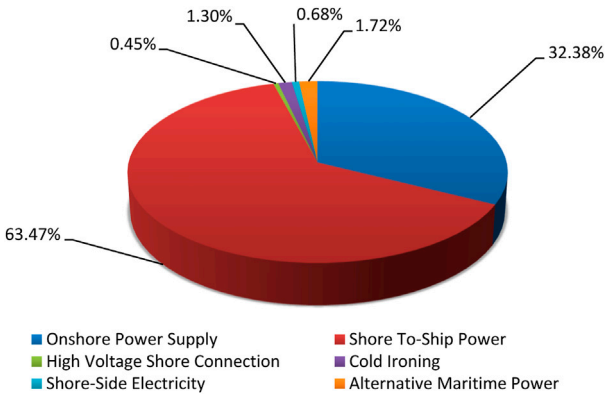


FIGURE 2
Percentage structure of Onshore Power System publications in the scientific databases analysed. Source: own study.

this stage, only 63 publications were qualified for deeper analysis, which accounted for 20.9% of those selected for analysis, which met the criteria of the 3rd stage (Table 1).

An important element in the considerations, and also the superior objective of the analysis, involves the issue of subject areas of research conducted in the area of OPS (Figure 4). As it was previously mentioned, OPS is the subject of interest to researchers from various fields and disciplines of knowledge. Some areas are excessively analysed in terms of research, while other, not much. Obviously, one can argue that perhaps these are the needs on the market, but one can also argue, quite the opposite, that perhaps these are topics convenient for researchers because they are consistent with their research interests, and not necessarily important for stakeholders.

A review of the scientific achievements to date related to the issue of OPS has shown that this issue is very topical and subject to analysis within different areas of knowledge, which can basically be divided into four main research areas, i.e.,: 1) environmental analysis (protection, pollution, emissions, etc.) 2) engineering analysis (technology, technical, etc.), 3) socio-economic analysis (financial, fiscal, business, management, etc.), and 4) legal analysis (legal standards, law regulations, etc.) (table 2).

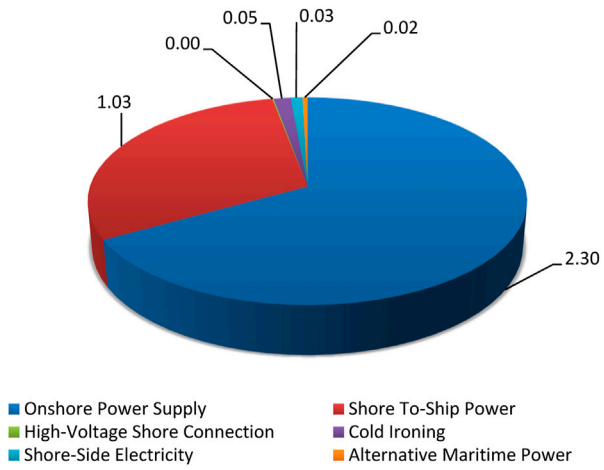


FIGURE 3
Number of publications related to Onshore Power System in the scientific databases analysed (in thousand). Source: own study.

The main reason for investing in OPS refers to the environmental benefits that reduce noise, vibrations and emissions of pollution from ships when they are berthed. Hence, in databases there are numerous publications on this subject presenting environmental benefits of using OPS technology (Kizielewicz and Skrzeszewska, 2021; Sciberras et al., 2015; Saxena et al., 2012; Kumar et al., 2018; He et al., 2020; Wang et al., 2021b; Reusser and Pérez, 2021; Chen, 2022; Siu et al., 2022; Su et al., 2023) in particular in the area of CO_2 reduction potential of OPS (Stolz et al., 2021). The research carried out by Reusser and Pérez (2021) (Wang et al., 2021b) is a good example showing "the reduction in the specific fuel consumption of the auxiliary engines, thus reducing emissions in 20% for CO_2, 34% for NO_x and 30% for SO_x" thanks to applying CI power configuration.

These results confirm the expediency of using CI and should constitute an important argument convincing to use these solutions. However, in the area of engineering research, there are numerous scientific works on the development of OPS infrastructure and various technologies in this respect used in seaports, including small and medium ones. Colarossi and Principi (2020) proposed the methodology of evaluation "the energy demand of ships at berth and producing electricity through a highly efficient cogeneration plant". They also studied the annual trend in electrical energy required by ships at berth and tried to estimate the capital, operational and maintenance costs. They presented the system which consists of a cogeneration plant that produces simultaneously heat to cover utilities in port area and power for ships at berth while (Colarossi and Principi, 2020). Numerous works also refer to the issue of connecting OPS to energy sources, including renewable sources. Interesting research on the use of wave energy extraction and wave energy converters was conducted by Cheng et al. (2022a) (Port Technology International, 2021; Cheng et al., 2022a). They described solutions combing floating breakwaters with wave energy converters and integrating them into a very large floating structure and developed mathematical models confirmed by experimental studies. It would be worth considering the use of offshore wave energy converters in OPS. There is still a gap in this area of research, and only a few experimental works have been done in this area. This could also be an interesting direction for future research.

Bearing in mind different energy supply needs of ships, due to their type and gross tonnage, research is also conducted on the electrical characteristics of OPS (Yang and Chai, 2016; Atallah et al., 2017; Gutierrez-Romero et al., 2019; Badakhshan et al., 2022; Farrukh et al., 2022). Bearing in mind different energy supply needs of ships, due to their type and gross tonnage, research is also conducted on the electrical characteristics of OPS (Bakar et al., 2022; Sciberras et al., 2015) and transient overvoltage protection of OPS System (Haddadian and Haddadian, 2011), and energy efficiency assessment of OPS (Karimi et al., 2022). A serious problem raised in scientific works also refers to the cost of energy transportation to OPS from power sources (Jung and Schindler, 2021), and also energy optimization of OPS systems. (Galkin and Tarnapowicz, 2022; Qiong and Xiao, 2015). Martínez-Lopez et al. (2021) noted that the environmental and economic benefits are only seen when OPS is powered by renewable sources.

The dominant group of works related to OPS also comprises studies in the area of socio-economic analysis, including mainly financial and management analysis of investments in OPS and benefits to local communities from OPS systems. Such analyses are a key element in the investment decision-making process of port authorities. Investments in OPS are highly capital-intensive, as already mentioned above, and many factors determine the success of this type of investment. Martínez-Lopez et al. (2021) presented "a calculation method to estimate a environmental charge in ports to incentivize cold Ironing use", which takes into account "a pollutant differentiation system by considering kinds of vessel, technical features, port localization and hinterlands populations". The team led by Najihah (2023) studied the "synergy of the cold ironing and microgrid system on ships" and also presented the "cold ironing cost analysis strategies from ports' and ships' points of view" (Najihah et al., 2023). It is one of the few scientific studies of this type containing both technological and economic aspects of CI. Many scientific papers refer to assessing the economic efficiency of investments in OPS including, *inter alia*, pricing strategy and sale mechanism of OPS (Stolz et al., 2021; Su et al., 2023) and cost-effective optimization analysis of OPS (Hulme, 2006).

The need for financial support for OPS from public funds (Wang et al., 2021a) is also raised, because of the environmental benefits for the region and local communities (Ballini and Bozzo, 2015). Moreover, a valuable source of information for stakeholders includes studies devoted to building strategy for OPS System (Yu et al., 2017) and control strategy for OPS (Ji et al., 2018) presenting also barriers and drivers to the implementation of OPS (Williamsson et al., 2022). Finally, there are also works related to energy management in seaports (Acciaro et al., 2014) including the management of OPS systems and optimization of daily use of OPS (Yu et al., 2022).

However, we can feel certain insufficiency when it comes to scientific studies in the area of managing shipboard energy (Sciberras et al., 2015), and the state of preparing ships to be equipped with devices adapted to connect to OPS system and to

TABLE 1 List of keywords that constitute grounds for searching the scientific databases.

Key phrases	N research papers and review papers in scientific databases available in open access form the period between 2014 and 2023						
1st stage	2nd Stage	3rd stage	4th stage	5th stage			6th stage
				EnA*	EgA*	SEA*	LA*	
Onshore AND Power AND Supply	Science Direct	n = 2295	n = 27	n = 9	n = 15	n = 1	n = 2	12
Shore AND "To-Ship" AND Power		n = 1034	n = 13	n = 3	n = 8	n = 2	n = 0	
"High-Voltage" AND "Shore Connection"		n = 4	n = 1	n = 0	n = 1	n = 0	n = 0	
"Cold Ironing"		n = 49	n = 12	n = 1	n = 10	n = 1	n = 0	
"Shore-Side" AND Electricity"		n = 34	n = 6	n = 0	n = 5	n = 1	n = 0	
Alternative AND "Maritime Power"		n = 15	n = 1	n = 0	n = 1	n = 0	n = 0	
TOTAL ScienceDirect		3432	60	13/12	40/68	5/15	2/3.4	
Onshore AND Power AND Supply	EBSCO	n = 21 511	n = 14	n = 2	n = 8	n = 2	n = 4	35
Shore AND "To-Ship" AND Power		n = 45 977	n = 45	n = 8	n = 30	n = 7	n = 0	
"High-Voltage" AND "Shore Connection"		n = 327	n = 26	n = 4	n = 18	n = 4	n = 0	
"Cold Ironing"		n = 910	n = 82	n = 14	n = 56	n = 12	n = 0	
"Shore-Side AND Electricity"		n = 420	n = 6	n = 1	n = 4	n = 1	n = 0	
Alternative AND "Maritime Power"		n = 1246	n = 3	n = 1	n = 2	n = 0	n = 0	
TOTAL EBSCO		70 424	176	30	118	24	4	
Onshore AND Power AND Supply	Web of Science	n = 194	n = 39	n = 7	n = 26	n = 6	n = 0	15
Shore AND "To-Ship" AND Power		n = 24	n = 18	n = 4	n = 12	n = 2	n = 0	
"High-Voltage" AND "Shore Connection"		n = 5	n = 0	n = 0	n = 0	n = 0	n = 0	
"Cold Ironing"		n = 1	n = 0	n = 0	n = 0	n = 0	n = 1	
"Shore-Side" AND Electricity"		n = 52	n = 6	n = 2	n = 4	n = 0	n = 0	
Alternative AND "Maritime Power"		n = 13	n = 2	n = 0	n = 2	n = 0	n = 0	
TOTAL Web of Science		279	65	13	44	8	0	
TOTAL Scientific bases		74 135	302	56	202	37	7	63
Percentage share in selected publications			100%	18,5%	66,9%	12,3%	2,3%	20,9%

EgA, Engineering analysis; EnA, Environmental analysis; SEA, Socio-economic analysis; LA, Legal analysis and documents.

ensure ship-to-grid integration (Vlachokostas et al., 2019). Few papers present the results of research on equipping the ships, by segments (cruise ships, ferry vessels, cargo, etc.) with devices enabling the connection to OPS, and future electricity and hydrogen demands for shipping (Ortiz-Imedio et al., 2021; Bakar et al., 2022). This is the area that should become the subject of future research. There is also a shortage of reliable economic analyses for shipowners regarding the benefits of using OPS.

We should also mention the legal and regulatory issues related to OPS systems, issued by the IMO (IMO, 1973) or IAPH (IAPH, 2010) and ISO (ISO/IEC/IEEE 80005-1:2012, 2012) regarding the technical aspects of OPS. These documents provide practical information about OPS, which constitute the subject of analyses and scientific discussions (IAPH, 2010; Yu et al., 2017; Tarnapowicz and German-Galkin, 2018).

Summing up the analysis of databases on previous research on OPS, it should be emphasized that the main streams of research concern the following areas.

- Environmental research area–the use of renewable energy sources to power the operational activities of seaports, including supplying ships with shore energy (OPS). The aim is to reduce greenhouse gas emissions and reduce the emission of harmful substances into the environment.
- Engineering research area–the use of modern technologies in the construction of OPS infrastructure, including in particular in relation to transformers, converters, energy storage devices in order to increase energy efficiency.
- The area of computer science research–the use of new technologies in the management of energy distribution in

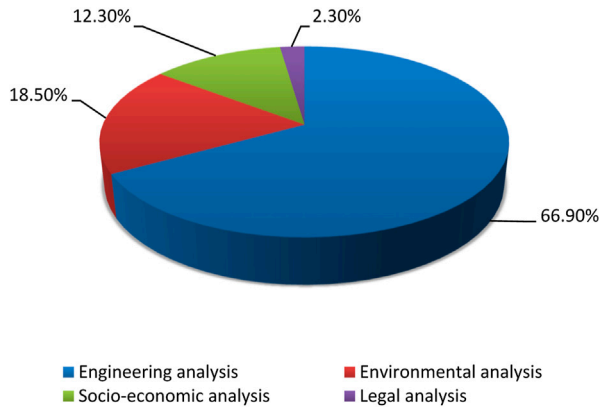


FIGURE 4
Subject areas of research publications in databases. Source: own study.

port areas in order to optimize its consumption, increase energy efficiency and estimate energy demand in the future.
- The area of economic research–using knowledge in the field of management and economics to prepare economic analyses of the efficiency of investments in OPS in relation to individual case studies.

It should be emphasized that the research conducted so far has focused mainly on the analysis of individual cases of engineering, organizational, technical or architectural solutions in the field of OPS. Stakeholders interested in investing in OPS and their development are looking for comprehensive analyses containing legal, economic, environmental and social analyses, showing a map of the benefits and costs resulting from investments in OPS. This type of interdisciplinary research would be more applicable and could have a greater popularizing value.

4.1 Conclusion and policy implications

In conclusion, it must be emphasized that despite numerous scientific studies dedicated to the economic analysis of investments in OPS, there is still a noticeable gap in studies directly dedicated to the management of investments in OPS and the social aspects of OPS. Considering that OPS are relatively new investments in seaports and new technological solutions for OPS continue to enter the market, port authorities are looking for the most effective solution models in this regard.

Frequently, port authorities also commission dedicated studies on the cost-effectiveness of investments and new solutions in OPS. However, these are industry-related studies rather than scientific papers and present a case study rather than a broader spectrum of solutions related to the solidity of investments in OPS.

At present, there are various technological solutions in the field of OPS, which are described above. The port authorities would be very interested not only in a comparative analysis of all these systems, in terms of energy efficiency, technological solutions and their development (engineering analysis), but also in the environmental benefits of specific OPS solutions (environmental analysis), and, most importantly, in assessing the economic viability of these investments (economic analysis). An important area of research on OPS is also the way of the OPS management, and development in seaports (business and management analysis).

One of the limitations of the analysis presented above involves the subjective criteria adopted by the author of this article. Limiting the analysis only to a decade, just only to research papers and review papers published between 2014 and 2023 and available as open access may affect the analysis results. Nevertheless, the conclusions that can be drawn present the general trends in research and areas of interest for scientists as regards the issue of OPS.

The analysis conducted in the study enables to indicate the areas of knowledge and research working currently on research related to OPS, which represented the main objective of this research. The analysis also proved that scientists use different abbreviations (e.g.,: OPS, SSP, CI, SSE, HSCV and AMP) and concepts to think and research the same phenomena. This can make analysis and search difficult and provides an important clue for researchers, It seems that the terminology in this area should be sorted out.

The review of the previous studies allowed to indicate the main directions for further scientific research to be conducted in this area of OPS, i.e.,.

- management of OPS systems and their integration with onshore energy systems,
- integration of offshore wave energy converters with onshore energy systems;
- budgeting policy of the investments in OPS co-financed by public funds;
- demonstrate the policy for economic efficiency of OPS for stakeholders (port authorities and ship owners);
- popularization of various technical and technological solutions related to OPS;
- social and environmental policy for OPS development;
- ships' equipment with a device enabling the connection to OPS;
- verification of the demand for OPS among shipowners;
- economic efficiency of OPS from the perspective of shipowners;
- identification of risks associated with investments in OPS.

Summing up, the future research should also address areas that are still insufficiently researched and described and represent gaps in research. The future research fields should concern.

- Comparative analyses of various engineering and technical solutions in the field of OPS, showing which of the solutions should be used under specific environmental conditions and financial capabilities of stakeholders.
- Economic analyses using an indicator analysis of the economic efficiency of individual OPS solutions together with a prospective analysis of the economic effects of these solutions in various variants for the region. Such analyses would be of great application value and could be an interesting source of information for entities managing port areas, including port authorities, local governments and private owners of the port lands.

TABLE 2 Research areas and topics in the field of onshore power supply.

Research areas	Research topics
Engineering analysis (technology, technical, etc.)	Technologies of OPS
	Optimal deployment of OPS
	Electrical characteristics of OPS
	Transient Overvoltage Protection of OPS System
	Energy Optimization of OPS systems
	Energy efficiency assessment of OPS
	Ship-to-grid integration
	Distance to power grids
	Fault diagnosis for onshore grid-connected converter in wind energy conversion systems
	Uninterruptable Shore-side Power Supply of OPS
	Implementing OPS from renewable energy sources
	Application Systems to Supply Onshore Grid Power to Offshore O&G Installations
	Potential of local solar generation for providing OPS
	Integrating Offshore Wind Farms with Unmanned Hydrogen and Battery Ships
	Installing cold ironing at small and medium ports
Socio-economic analysis (financial, fiscal, business, management, etc.)	Cooperative optimization of OPS
	Barriers and Drivers to the Implementation of OPS
	Strategy for OPS System
	Control Strategy for OPS
	Energy management in seaports
	Government subsidize for OPS
	Electricity subsidy efficiency of OPS
	Economic Benefits of OPS
	Cost-effective optimization analysis of OPS
	Optimization of daily use of OPS
	Sale Mechanism of Onshore Power Supply
	Pricing Strategy of Cold Ironing
	Improving OPS project economics
	Techno-economic analysis of OPS
	Socio-economic benefit of cold-ironing technology
	Managing Shipboard Energy
	Ship berthing forecasting for cold ironing
	Future electricity and hydrogen demands for shipping
Environmental analysis (protection, pollution, emissions, etc.)	Environmental Benefits of Using low-Sulphur Oil and OPS Technology
	Evaluation of the emission impact of OPS
	Environmental emission Impact of OPS
	CO2 reduction potential of OPS
	Environmental charges to boost Cold Ironing

(Continued on following page)

TABLE 2 (*Continued*) Research areas and topics in the field of onshore power supply.

Research areas	Research topics
	Reducing shipboard emissions
	Environmental Ben46efits of OPS
Legal analysis (legal standards, law regulations, etc.)	Regulation strategy for OPS
	Standardization in the Design of OPS

Source: own elaboration on the base of: (IMO, 2015; Innes and Monios, 2018; IAPH, 2010; Winkel et al., 2016; Stolz et al., 2021; Bakar et al., 2022; Bullock et al., 2023; Hulme, 2006; AAPA, 2007; Haddadian and Haddadian, 2011; Acciaro et al., 2014; Qiong and Xiao, 2015; Sciberras et al., 2015; Yang and Chai, 2016; Atallah et al., 2017; Yu et al., 2017; Chengdi et al., 2018; He et al., 2018; Ji et al., 2018; Tarnapowicz and German-Galkin, 2018; Gutierrez-Romero et al., 2019; Kumar et al., 2019; Sun et al., 2019; Vlachokostas et al., 2019; Zis, 2019; Colarossi and Principi, 2020; He et al., 2020; Wang et al., 2020; Wang et al., 2021a; Jung and Schindler, 2021; Kizielewicz and Skrzeszewska, 2021; Martínez-Lopez et al., 2021; Ortiz-Imedio et al., 2021; Peng et al., 2021; Spengler and Tovar, 2021; Badakhshan et al., 2022; Farrukh et al., 2022; Galkin and Tarnapowicz, 2022; Karimi et al., 2022; Qiu et al., 2022; Williamsson et al., 2022; Yu et al., 2022; Najihah et al., 2023; Tarnapowicz and German-Galkin, 2018; Port Technology International, 2021; Reusser and Pérez, 2021; Cheng et al., 2022a; Cheng et al., 2022b; Kizielewicz, 2023).

- Environmental analyses showing the benefits of using various OPS solutions and, on the other hand, the benefits and costs associated with not using these solutions in ports for the region and the environment.
- Engineering analyses showing in a comprehensive way the various possibilities of supplying OPS with energy obtained from renewable energy sources. Especially in coastal areas, there are opportunities to use water energy, wind energy and solar energy. Although such studies are already carried out on a random basis, they are individual case studies and not comprehensive and variant analyses. In Europe, investments in the area of offshore wind farms are currently intensifying. The costs of preparing and maintaining the OPS supply infrastructure are significant factors withholding the development of OPS. It is crucial to demonstrate the benefits of using renewable energy sources to power OPS in order to persuade decision-makers to invest in OPS.
- A holistic and interdisciplinary approach showing the state and prospects for the development of OPS in port areas and opportunities for regions thanks to investments in OPS, as well as the risks for the region in the event of a departure from investments in this area.
- Presentation of the state of technical readiness of the fleet of vessels in individual segments (cruies ships, ferry vessels, cargo ships, etc.) for the use of OPS, together with an analysis of plans for the installation of devices enabling connection to shore power supply. Currently, analyses in this area are fragmentary, and the lack of information in this area does not support investors' decisions and causes great concern.
- Develop examples of good practice in relation to the development of international standards and guidelines for the design, installation and operation of onshore energy infrastructure.
- Social analyses showing what is the attitude of stakeholders in relation to both OPS and renewable energy sources. When talking about stakeholders, we must not forget about the inhabitants of port cities, who are not always enthusiastic about, for example, the construction of wind farms and are not aware of the benefits of investments in this area.

So far, the analysis of research results shows that seaport authorities are open to create the policy of investments development supporting the environment management, facilitating the environmental improvement and reducing the amount of pollutants emitted to the environment (Kizielewicz, 2023), but they need support in the form of scientific analyses on the best technological solutions, models, and sources of financing for the investments in OPS. The challenges faced by seaports, especially in Europe, resulting from climate change and restrictive EU and UN regulations, mean that port authorities will have no choice but to invest in this area.

In future research, a set of other phrases can be reviewed, and other publications can be selected for analysis. A significant limitation certainly involves including in the scope of research the publications with limited access which require paid access thereto. This type of analysis could be very expensive and difficult to implement, but perhaps it would provide a different picture of the ongoing scientific research in the area of OPS.

Author contributions

JK: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, 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 research was fully funded by a grant WZNJ/2024/PZ/09.

Conflict of interest

The 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.

References

AAPA (2007). "Use of shore-side power for ocean-going vessels," in *White paper, American Association of Port Authorities* (Tetra Tech, Inc), 1035.

Acciaro, M., Ghiara, H., and Cusano, M. I. (2014). Energy management in seaports: a new role for port authorities. *Energy Policy* 71, 4–12. doi:10.1016/j.enpol.2014.04.013

Atallah, A., Kerin, U., and Audring, D. (2017). *Examining the operational and environmental benefits of using high-voltage AC transmission systems to supply onshore grid power to offshore O&G installations*.

Badakhshan, S., Senemmar, S., Li, H., and Zhang, J. (2022). "Integrating offshore wind farms with unmanned hydrogen and battery ships," in IEEE Kansas Power and Energy Conference, Manhattan, Manhattan, KS, USA, April 25 and 26, 2022, 1–6. doi:10.1109/KPEC54747.2022.9814722

Bakar, N. N. A., Bazmohammadi, N., Çimen, H., Uyanik, T., Vasquez, J. C., and Guerrero, J. M. (2022). Data-driven ship berthing forecasting for cold ironing in maritime transportation. *Appl. Energy* 326, 1–12. 119947. doi:10.1016/j.apenergy.2022.119947

Ballini, F., and Bozzo, R. (2015). Air pollution from ships in ports: the socio-economic benefit of cold-ironing technology. *Res. Transp. Bus. Manag.* 17, 92–98. doi:10.1016/j.rtbm.2015.10.007

Bouman, E. A., Lindstad, E., Rialland, A. I., and Strømman, A. H. (2017). State-of-the-art technologies, measures, and potential for reducing GHG emissions from shipping – a review. *Transp. Res. Part D* 52, 408–421. doi:10.1016/j.trd.2017.03.022

Bullock, S., Higgins, E., Crossan, J., and Larkin, A. (2023). *Marine policy 152 (2023) improving shore power project economics at the port of aberdeen*. Amsterdam, Netherlands: Elsevier, 2–9.

Chen, C. J., Zheng, T., Garg, A., Xu, L., Li, S., and Fei, Y. (2019). Alternative Maritime Power application as a green port strategy: barriers in China. *Barriers J. Clean. Prod.* 213, 825–837. doi:10.1016/j.jclepro.2018.12.177

Chen, G. (2022). Pinning control of complex dynamical networks. *IEEE Trans. Consumer Electron.* 68 (4), 336–343. doi:10.1109/TCE.2022.3200488

Cheng, Y., Chen, Xi, Dai, S. S., Ji, C. Y., Collu, M., Li, M., et al. (2022a). Wave Energy extraction and hydroelastic response reduction of modular floating breakwaters as array wave energy converters integrated into a very large floating structure. *Appl. Energy* 306 (117953), 1–32. doi:10.1016/j.coastaleng.2022.104188

Cheng, Y., Fu, L., Dai, S. S., Collu, M., Ji, C. Y., Yuan, Z. M., et al. (2022b). Experimental and numerical investigation of WEC-type floating breakwaters: a single-pontoon oscillating buoy and a dual-pontoon oscillating water column. *Coast. Eng.* 17, 1–16. doi:10.1016/j.coastaleng.2022.104188

Chengdi, D., Yan, D., Zhongyi, L., Liu, S., and Jia, S. (2018). "Study on the power supply and sale mechanism of onshore power supply under the background of electric power system reform," in 2018 China International Conference on Electricity Distribution. IEEE, Tianjin, China, 17-19 September 2018.

Colarossi, D., and Principi, P. (2020). Technical analysis and economic evaluation of a complex shore-to-ship power supply system. *Appl. Therm. Eng.* 181, 1–11. doi:10.1016/j.applthermaleng.2020.115988

EC (2019). *Communication from the commission, the European green deal, COM(2019) 640 final*, 1–24.

Farrukh, F., Dunks, C., and Marie, M. (2022). "Assessment of the potential of local solar generation for providing ship shore power in the Norwegian harbour Port of Borg," in Conference: 18th International Conference on the European Energy Market, Ljubljana, Slovenia, 13-15 September 2022, 1–6. doi:10.1109/EEM54602.2022.9921031

Galkin, S., and Tarnapowicz, D. (2022). Energy optimization of the 'shore to ship' system-A universal power system for ships at berth in a port. *Sensors* 20 (14), 1–20. doi:10.3390/s20143815

General Assembly (2015). *UN, transforming our world: the 2030, Agenda for sustainable development, A/RES/70/1*, 5–38.

Gutierrez-Romero, J. E., Jose, E., Esteve-Perez, J., and Zamora, B. (2019). Implementing Onshore Power Supply from renewable energy sources for requirements of ships at berth. *Appl. Energy* 255, 113883. doi:10.1016/j.apenergy.2019.113883

Haddadian, P. D., and Haddadian, V. (2011). Transient overvoltage protection of shore-to-ship power supply system. *IEEE Trans. Industry Appl.* 47 (3), 1193–1200. doi:10.1109/tia.2011.2125772

He, J., Li, X., Xu, H., Zhu, J., Dai, P., and Chu, H. (2018). "Review and discussion on standards for shore-to-ship power supply system," in International conference on material engineering and application (ICMEA 2017). Vol. 146, Wuhan, China, December 15th-17th, 2017, 22–26.

He, W., Chen, R., Li, T., Hu, B., Tian, Y., and Meng, W. (2020). Comparative study on environmental benefits of using low-sulphur oil and shore power technology for ship berthing. *E3S Web Conf.* 145, 02010–02015. doi:10.1051/e3sconf/202014502010

Hulme, C. T. (2006). Using cost effectiveness analysis; a beginners guide. *Evid. Based Libr. Inf. Pract.* 1 (4), 17–29. doi:10.18438/B81S34

IAPH (2010). *Onshore power supply*. Available at: https://sustainableworldports.org/project/iaph-onshore-power-supply.

IMO (1973). *Final act of the international conference on maritime pollution, date of access: 23/04/2022*. Available at: https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/ConferencesMeetings/Documents/MARPOL%201973 (accessed on April 17, 2022).

IMO (2015). Guidelines for exhaust gas cleaning systems, marine environment protection committee. *IMO Publ. Documents Resolut. MEPC* 259 (68), 1–23.

Innes, A., and Monios, J. (2018). Identifying the unique challenges of installing cold ironing at small and medium ports – the case of aberdeen. *Transp. Res. Part D Transp. Environ.* 62, 298–313. doi:10.1016/j.trd.2018.02.004

International Maritime Organization IMO (2012). *On-shore power supply, MEPC.1/Circ.794 9 october 2012*. Available at: https://wwwcdn.imo.org/localresources/en/OurWork/Environment/Documents/Circ.794.pdf.

ISO/IEC/IEEE 80005-1:2012 (2012). *Utility connections in port — Part 1: high voltage shore connection (HVSC) systems — general requirements*. Available at: https://www.iso.org/standard/53588.html.

Ji, Z., Zhao, Z., Wang, J., and Lv, Z. (2018). A flexible control strategy for shore-to-ship power system in terms of grid-connected and off-grid switch. *Pol. Marit. Res.* 25, 139–148. doi:10.2478/pomr-2018-0085

Jung, C., and Schindler, D. (2021). Distance to power grids and consideration criteria reduce global wind energy potential the most. *J. Clean. Prod.* 317, 128472–128512. doi:10.1016/j.jclepro.2021.128472

Karimi, S., Zadeh, M., and Suul, J. A. (2022). A multi-layer framework for energy efficiency assessment of shore-to-ship fast charging systems including onshore batteries. *IET Electr. Syst. Transp.* 12 (4), 269–286. doi:10.1049/els2.12052

Kizielewicz, J. (2023). Environmentally friendly cruise seaports in northern Europe – onshore power supply. *Int. J. Mar. Navigation Saf. Sea Transp.* 17 (2), 357–363. doi:10.12716/1001.17.02.12

Kizielewicz, J., and Skrzeszewska, K. (2021). Identifying actions to prepare electricity infrastructure in seaports for future power supplying cruise ships with energy from land. *Energies* 14 (23), 8173. No 8173. doi:10.3390/en14238173

Kumar, J., Kumpulainen, L., and Kauhaniemi, K. (2019). Technical design aspects of harbour area grid for shore to ship power: state of the art and future solutions. *Int. J. Electr. Power Energy Syst.* 104, 840–852. doi:10.1016/j.ijepes.2018.07.051

Kumar, N., Singh, B., and Panigrahi, B. K. (2018). *Grid synchronisation framework for partially shaded solar PV-based microgrid using intelligent control strategy*, 829–837. doi:10.1049/iet-gtd.2018.6079

Li, K., and Du, K. (2020). Research on onshore power supply system in port for ships. *Earth Environ. Sci.* 558, 052022–052025. doi:10.1088/1755-1315/558/5/052022

Martínez-Lopez, A., Romero-Filgueira, A., and Chica, M. (2021). Specific environmental charges to boost cold ironing use in the European short sea shipping transportation research Part D. *Transp. Environ.* 94, 1–14. doi:10.1016/j.trd.2021.102775

Mattsson, N., Verendel, V., Hedenus, F., and Reichenberg, L. (2021). *An autopilot for energy models – automatic generation of renewable supply curves, hourly capacity factors and hourly synthetic electricity demand for arbitrary world regions Energy Strategy Reviews 33*. Amsterdam, Netherlands: Elsevier, 2–9.

Najihah, N., Bakar, A., Bazmohammadi, B., Vasquez, J. C., and Guerrero, J. M. (2023). Electrification of onshore power systems in maritime transportation towards decarbonization of ports: a review of the cold ironing technology. *Renew. Sustain. Energy Rev.* 178, 1–16. doi:10.1016/j.rser.2023.113243

OPS Master Plan (2021). *OPS master plan for Spanish ports project*. Available at: http://poweratberth.eu/?page_id=219&lang=en.

Ortiz-Imedio, R., Caglayan, D. G., Ortiz, A., Heinrichs, H., Robinius, M., Stolten, D., et al. (2021). Power-to-Ships: future electricity and hydrogen demands for shipping on the Atlantic coast of Europe in 2050. *Energy* 228, 120660–120727. doi:10.1016/j.energy.2021.120660

Peng, Y., Dong, M., Li, X. D., Liu, H. K., and Wang, W. Y. (2021). Cooperative optimization of shore power allocation and berth allocation: a balance between cost and environmental benefit. *J. Clean. Prod.* 279, 123816–123818. doi:10.1016/j.jclepro.2020.123816

Port Technology International (2021). What is onshore power? Available at: http://www.porttechnology.org/news/what-is-onshore-power (Accessed May 7, 2023).

Puig, M., Azarkamand, S., Wooldridge, C., Selén, V., and Darbra, R. M. (2022). *Insights on the environmental management system of the European port sector 150550, Science of the Total Environment, 806*. Amsterdam, Netherlands: Elsevier, 2–12.

Qi, J., Wanga, S., and Peng, C. (2020). *Shore power management for maritime transportation: status and perspectives Maritime Transport Research, (1)*. Amsterdam, Netherlands: Elsevier, 1–11.

Qiong, S. Y., and Xiao, L. M. (2015). "Technology of uninterruptable shore-side power supply for berthing vessels and its application," in Proceedings of The 3rd International Conference On Advances, Energy and Environmental Science AER-Advances in Engineering Research, Vol. 31, Zhuhai, China, July 25-26, 2015, 371–375.

Qiu, J., Tao, Y., Lai, S., and Zhao, J. (2022). Pricing strategy of cold ironing services for all-electric ships based on carbon integrated electricity price. *IEEE Trans. Sustain. Energy IEEE Trans. Sustain. Energy Sustain. Energy, IEEE Trans.*13 (3), 1553–1565. doi:10.1109/tste.2022.3157645

Reusser, C. A., and Pérez, J. R. (2021). Evaluation of the emission impact of cold-ironing power systems, using a Bi-directional power flow control strategy. *Sustainability* 13 (1), 334. doi:10.3390/su13010334

Saxena, V., Kumar, N. N., Singh, B., and Panigrahi, B. K. (2012). 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, 1469–1478. doi:10.1109/TEC.2021.3059754

Sciberras, E. A., Zahawi, B., and Atkinson, D. J. (2015). Electrical characteristics of cold ironing energy supply for berthed ships. *Transp. Res. Part D-Transport Environ.* 39, 31–43. doi:10.1016/j.trd.2015.05.007

Siu, J. Y., Kumar, N., and Panda, S. K. (2022). Command authentication using multiagent system for attacks on the economic dispatch problem. *IEEE Trans. Industry Appl.* 58 (4), 4381–4393. doi:10.1109/TIA.2022.3172240

Spengler, T., and Tovar, B. (2021). Potential of cold-ironing for the reduction of externalities from in-port shipping emissions: the state-owned Spanish port system case. *J. Environ. Manag.* 279, 111807. doi:10.1016/j.jenvman.2020.111807

Stolz, B., Held, M., Georges, G., and Boulouchos, K. (2021). The CO2 reduction potential of shore-side electricity in Europe. *Appl. Energy* 285, 116425–116512. doi:10.1016/j.apenergy.2020.116425

Su, S., Li, W., Mou, J., Garg, A. R., Gao, L., and Liu, J. (2023). A hybrid battery equivalent circuit model, deep learning, and transfer learning for battery state monitoring. *IEEE Trans. Transp. Electrification* 9, 1113–1127. doi:10.1109/tte.2022.3204843

Sun, Y., Zhu, L., Xu, Z., Xiao, L., Zhang, J., and Zhang, J. (2019). Characteristic analysis and forecast of electricity supply and demand in APEC. *Glob. Energy Interconnect. V.* 2 (6), 413–422. doi:10.1016/j.gloei.2019.11.016

Tarnapowicz, D., and German-Galkin, S. (2018). International standardization in the design of "shore to ship" - power supply systems of ships in port. *Manag. Syst. Prod. Eng.* 26 (1), 9–13. doi:10.2478/mspe-2018-0001

The Port of Bremen (2023). Cruise Terminal, modern, secure, purpose-built. Available at: https://www.bremenports.de/en/ports/bremerhaven.

Vaishnav, P., Fischbeck, P. S., Morgan, M. G., and Corbett, J. J. (2016). Shore power for vessels calling at U.S. Ports: benefits and costs. *Environ Sci. Technol.* 50 (3), 1102–1110. doi:10.1021/acs.est.5b04860

Viana, M., Hammingh, P., Colette, A., Querol, X., Degraeuwe, B., Vlieger, I., et al. (2014). Impact of maritime transport emissions on coastal air quality in Europe Atmospheric Environment. *Atmos. Environ.* 90, 1–9. 96e105. doi:10.1016/j.atmosenv.2014.03.046

Vlachokostas, A., Prousalidis, J., Spathis, D., Nikitas, M., Kourmpelis, K., Dallas, S., et al. (2019). "Ship-to-grid integration: environmental mitigation and critical infrastructure resilience," in 2019 IEEE Electric Ship Technologies Symposium (ESTS), Washington, DC, 542–547. doi:10.1109/ESTS.2019.8847858

Wang, L. F., Liang, C. J., Shi, J., Molavi, A., Lim, G., and Zhang, Y. (2021a). A bilevel hybrid economic approach for optimal deployment of onshore power supply in maritime ports. *Energy* 292, 264–274. doi:10.1016/j.apenergy.2021.116892

Wang, X. Y., Chen, Y. Z., Liu, Z. Q., Zhang, H. W., Li, Q. Y., Wang, J. X., et al. (2021b). High-density memristor-CMOS ternary logic family. *IEEE Trans. Circuits Syst. I Regul. Pap.* 68 (1), 264–274. doi:10.1109/TCSI.2020.3027693

Wang, Y., Ding, W., Dai, L., and Jing, D. (2020). How would government subsidize the port on shore side electricity usage improvement? *J. Clean. Prod.* 278 (7), 123893–123896. doi:10.1016/j.jclepro.2020.123893

Williamsson, J., Costa, N., Santen, V., and Rogerson, S. (2022). Barriers and drivers to the implementation of onshore power supply-A literature review. *Sustainability* 14 (10), 2–16. doi:10.3390/su14106072

Winkel, R., Weddige, U., Johnsen, D., Hoen, V., and Papaefthimiou, S. (2016). Shore side electricity in Europe: potential and environmental benefits. *Energy Policy* 88, 584–593. doi:10.1016/j.enpol.2015.07.013

World Port Sustainability Program (2021). *Hamburg port authority – onshore power supply*. Available at: https://sustainableworldports.org//project//port-of-hamburgonshore-power-supply-ops//.

Yang, Z. M., and Chai, Y. (2016). A survey of fault diagnosis for onshore grid-connected converter in wind energy conversion systems. *Renew. Sustain. Energy Rev.* 66, 345–359. doi:10.1016/j.rser.2016.08.006

Yu, J. J., Voss, S., and Song, X. Q. (2022). Multi-objective optimization of daily use of shore side electricity integrated with quayside operation. *J. Clean. Prod.* 351, 1–7. doi:10.1016/j.jclepro.2022.131406

Yu, M. D., Huang, W. T., Tai, N. L., Zheng, X. D., Ma, Z. J., and Wang, Y. (2017). Advanced microgrid and its multi-objective regulation strategy for shore supply. *J. Eng.* 2017 (13), 1590–1594. doi:10.1049/joe.2017.0600

Zhang, H., Tomasgard, A., rage Rugstad Knudsen, B. R., Svendsen, H. G., Bakker, S. J., and Grossmann, I. E. (2022). *Model. analysis offshore energy hubs Energy* 261, 1–17. Part A. doi:10.1016/j.energy.2022.125219

Zis, T. P. V. (2019). Prospects of cold ironing as an emissions reduction option. *Transp. Res. Part A* 119, 82–95. doi:10.1016/j.tra.2018.11.003

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 02 April 2024
DOI 10.3389/fenrg.2024.1377161

Check for updates

OPEN ACCESS

EDITED BY
Yang Yang,
Nanjing University of Posts and Telecommunications, China

REVIEWED BY
Zhengmao Li,
Aalto University, Finland
Yunyang Zou,
Nanyang Technological University, Singapore
Chuanshen Wu,
Cardiff University, United Kingdom
Qiang Xing,
Nanjing University of Posts and Telecommunications, China

*CORRESPONDENCE
Peng Li,
lipeng111225@163.com

RECEIVED 26 January 2024
ACCEPTED 05 March 2024
PUBLISHED 02 April 2024

CITATION
Li P, Zhao R, Feng H, Wang H and Yu Z (2024), Channel prediction method based on the data-driving for distribution automation main station. *Front. Energy Res.* 12:1377161. doi: 10.3389/fenrg.2024.1377161

COPYRIGHT
© 2024 Li, Zhao, Feng, Wang and Yu. 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.

Channel prediction method based on the data-driving for distribution automation main station

Peng Li[1,2]*, Ruifeng Zhao[3], Huihui Feng[1,2], Hailong Wang[1,2] and Zhiwen Yu[3]

[1]NARI Group Co., Ltd, Nanjing, China, [2]NARI-TECH Nanjing Control Systems Co., Ltd, Nanjing, China, [3]Power Dispatching Control Center, Co., Ltd., Guangzhou, China

A data-driven channel prediction method for distribution automation master is proposed to address the poor quality of communication network and communication system transmission problems in distribution network communication. In this paper, an adaptive broad learning network (ABLN) consisting of a standard broad learning network and a hybrid learning network is introduced to predict the channel state information of the communication system. Among them, the hybrid learning network is used to solve the ill-conditioned solution problem when estimating the output weight matrix of the standard broad learning network. Therefore, the ABLN produces sparse output weight matrices and provides excellent prediction performance. In the simulation analysis, the outdoor and indoor scenes are considered based on OFDM system. The prediction performance of ABLN is subsequently evaluated in one step prediction and multistep prediction. The results show that for the prediction performance is concerned, the maximum improvement of ABLN is about 96.49% as compared to other evaluation models, indicating that the CSI is effectively predicted by the ABLN to support the adaptive transmission of the main station of the distribution automation and to satisfy the quality of the communication network of the distribution network.

KEYWORDS

distribution automation main station, channel prediction, data-driving, learning network, broad learning

1 Introduction

Currently, the distribution automation main station (DAMS) tends to huge network size, wide distribution, harsh environment, and frequent changes of distribution sites (Gu, 2017). The distribution communication network, as the nervous system of distribution network automation, is bound to meet the high-quality communication needs. Therefore, high-quality distribution network communication system is the foundation of the smart distribution network construction. The distribution network automation construction presents many challenges to the wireless communication system (Pan L et al., 2023). Generally speaking, the receiver of wireless communication system needs to estimate the channel state information (CSI) of transmission environment through channel estimation technique, and feedback the CSI to the transmitter. However, in fast time-varying communication environments, the CSI

tends to be outdated, which degrades the adaptive transmission performance. Therefore, the channel prediction based on the outdated CSI is important to support the adaptive transmission performance of wireless communication systems such as power IoT (by Flam et al., 2006).

Currently, channel prediction methods can be classified into three categories, i.e., linear prediction methods parameter class prediction methods and nonlinear prediction methods (Sarankumar R et al., 2016) Among them, linear prediction methods are mainly used to predict the next channel state information sample by linearly weighting the sum of several CSI samples in the past by linear fitting. Currently, channel prediction methods can be classified into three categories,i.e., linear prediction methods parameter class prediction methods and nonlinear prediction methods. Among them, linear prediction methods are mainly used to predict the value of the next channel state information sampling point by linearly fitting the past several channel state information sampling points, i.e., by linearly weighting the sum. Linear prediction methods include the auto-regression (AR), the recursive least squares (RLS) and the affine projection algorithm (APA) (Kapoor et al., 2018). The parameter prediction method mainly estimates those relevant parameters of transmission delay, such as the power, the delay and the Doppler shift. In addition, errors in the estimated channel statistical characteristics and parameters reduce the accuracy of the channel prediction, so the errors in the parameter class prediction need to be reduced in order to improve the performance of adaptive transmission techniques for wireless communication systems. Scholars have developed relevant studies on parameter class channel prediction methods. Such as, Niu G Q et al. (2014) and Yang et al. (2021) proposed a new predictor that utilizes the powerful time series prediction capability of deep learning. The prediction result indicates that the deep learning can offer significant performance improvement compared with the traditional predictor. To overcome the problem caused by the nonlinearity of the transmission channel and inter-symbol interference caused by multipath effects, Tan and Wang, (2023) used deep learning networks to optimally research the channel equalization for the optical fiber communication networks, and constructed an optical fiber communication network channel model, and used deep learning networks to estimate the communication network channel loads and operational states. Trivedi and Kumar, (2018) used a scheduler based on a standardized SNR for selecting users for data transmission, the scheduler has higher data rate and long-range transmission capabilities without requiring much power or bandwidth. In addition, this paper presents a comparative assessment of the bit error rate (BER) performance of multi-user multiple input multiple output orthogonal frequency division multiplexing (MU-MIMO-OFDM) and MU-MIMO single-carrier frequency-division multiple access (MU-MIMO-SCFDMA) and investigates the impact of various factors, e.g., the CSI imperfections, network heterogeneity, and other factors on the communication transmission; Bai et al. (2020) proposed a prediction method based on the long short term memory (LSTM) network and developed an incremental learning scheme for dynamic scenarios, which makes the LSTM predictor run online; (Multiple-Input Multiple-Output, MIMO) system, based on the measured data of 2.35 GHz band in the road-wall scenario, C Xue et al. (2021) proposes a Convolutional Long Shore-Term Memory (CLSTM) and CNN combination of Conv- CLSTM channel prediction model for typical channel state information, Les K-factor, RMS delay extension and angle extension characteristics prediction. Son and Han, (2021) proposed the channel adaptive transmission (CAT), which uses the LSTM network for channel prediction and the prediction accuracy is over 97%, indicating that this algorithm can be effectively used for channel prediction. Jiang and Schotten, (2019) used the recurrent neural network (RNN) to construct a frequency domain channel predictor, which was integrated into the MIMO-OFDM system to improve the correctness of antenna selection. To avoid the degradation of communication quality, Ding and Hirose, (2013) proposed a high-accuracy time-varying channel prediction by using channel prediction with linear extrapolation and Varangian extrapolation of frequency-domain parameters, and by combining a multilayered complex neural network (CVNN) with a linear FM permutation method. The proposed CVNN-based predictive channel prediction accuracy is experimentally proved to be better than the traditional prediction methods (Ding and Hirose, 2014). Xu and Han, (2016) proposed a new model adaptive elastic echo state network (ESN), which adopts the adaptive elastic network algorithm to compute the unknown weights and combines the advantages of quadratic regularization and adaptive weighted lasso contraction to deal with the covariance problem.

In recent years, the deep learning has been successfully applied in many fields and played an important role in the field of the artificial intelligence field (Schofield et al., 2019). Wang et al. (2019b) proposed a hybrid deep learning method of Convolutional Neural Network (CNN) and LSTM to get the CSI of downlink channel based on the CSI of uplink channel in FDD system. By extending an LSTM for reconstructing CSI, Wang et al. (2019a) proposed a real-time CSI feedback framework applied to point-to-point massive MIMO. The application of deep learning in channel prediction successfully solves the problem of traditional manual methods that are overly dependent on channel-specific parameters and has a stronger ability to adapt to the environment. However, the deep learning methods still have some drawbacks. On the one hand, when facing the high-dimensional data, the deep learning is usually trained with a complex structure, which means that many model parameters need to be adjusted. On the other hand, when some new samples are added, the deep learning often needs to re-train the model, which is a quite time-consuming process. To address the issue above, Chen et al. (2018) developed the broad learning system (BLS). Unlike the deep learning model, the BLS only has two horizontally aligned hidden layers, i.e., the feature layer and the enhancement layer (Suganthan and Katuwal, 2021; Pao and Takefuji, 2021). In the training process of the BLS model, the input data are firstly generated into feature nodes by feature mapping, and then the feature nodes are enhanced into augmentation nodes by nonlinear changes. Finally, the output of the feature layer and the output of the enhancement layer are connected to generate the result of the final output layer. The output weights of the final output layer can be obtained quickly using the ridge regression algorithm without complex computation (Chen and Liu, 2018). In addition, BLS has an attractive advantage of

the universal approximation property (Jin and Chen, 2018). In particular, the incremental learning algorithms of the BLS can be quickly reshaped without the need for complete retraining from scratch.

With these solid foundations mentioned above, a large amount of work on BLS has been reported. Kurtz, (1987) incorporated the and paradigm combination into the regular term of the elastic network into the BLS model thus obtaining ENBLS. ENBLS can benefit from the trend of both ridge regression and sparse solutions. However, there are still many drawbacks of the shallow structure, and the expressive ability is significantly weaker than that of the deep structure. For this reason, many scholars started to improve the standard BLS. Then, many more complex BLS appeared. Such as, Feng and Chen, (2020) integrated BLS with Takagi-Sugeno fuzzy system to generate the FBLS. In addition, to extend the applicability of BLS, Zhao et al. (2020) obtained Semi-Supervised BLS (SS-BLS) by introducing popular learning and thus extending BLS to semi-supervised learning. However, the SS-BLS requires all labelled training datasets, which limits the practicability in practical scenarios. Another structure of online semi-supervised BLS (Online SS-BLS, OSSBLS) was investigated in Pu and Li, (2020). In addition, Min et al. (2019) improved the BLS and obtained a Structured Manifold BLS (SM-BLS), which was used to predict the time-series. Zhang et al. (2020) also applied the BLS for the emotion recognition. Then, Liang, (2023) investigated a class of the distributed learning algorithms based on the BLS, which used a quantization strategy to reduce the number of bits per transmission and a communication review strategy to reduce the total number of transmissions and minimize the communication cost (Huong et al., 2023).

To meet the demand of the information transmission quality of the DAMS and solve the problem that the expired CSI reduces the adaptive transmission performance of wireless communication system, this paper proposes a channel prediction method for ADSM based on the ABLN, including the standard BLN and the hybrid regularization network. Thereinto, the ill-conditioned solution of the output weight matrix of the former is mainly solved by the latter. The hybrid regularization network contains two layers, the first layer is an adaptive weight factor generation network based on regularization, and the second layer is the output weight matrix estimation network based on the elastic network. The hybrid regularization network has the oracle property, which can effectively solve the output weight ill-conditioned solution problem of the BLN. Thus, the ABLN can provide well channel prediction performance.

2 Revelant theory

2.1 Channel estimation for OFDM systems

The channel prediction is an important technique to support the adaptive transmission of power IoT communication systems such as distribution network automation master stations. In this paper, it is assumed that the SCI changes slowly or steadily in a frame time, so the channel estimation process can be described as follows. Let $s_i(n)$ denote the n-th sample of the i-th complex-valued baseband time-domain transmission signal, then $s_i(n)$ can be expressed as

$$s_i(n) = \sum_{l=0}^{L-1} \boldsymbol{S}_i(l) e^{j2\pi nl/L} \tag{1}$$

where $\boldsymbol{S}_i(l)$ denotes the frequency-domain transmit signal on the lth subcarrier of the i-th pilot OFDM symbol, l = 0, 1, 2, ... , $L-1$, L is the total number of subcarriers in each OFDM symbol. When the transmitting antenna sends out the transmit signal, the transmit signal reaches the receiving antenna through the time domain channel. Therefore, the receiving end estimates the target CSI by analyzing the received signal. Currently, the two commonly used channel estimation methods are least squares (LS) method and the minimum mean square error (MMSE) method. The process of the channel estimation by LS method is [29]

$$Y_i(m) = \frac{E_i(m)}{S_i(m)} + R_i(m) = Y_i(m) + R_i(m) \tag{2}$$

where $Y_i(m)$, $Y_i(m)$, $E_i(m)$ and $R_i(m)$ are the estimated CSI of the i-th subcarrier of the OFDM symbol, the ideal CSI, the received symbols, and the corresponding noise for the first subcarrier of the OFDM symbol at the first pilot subcarrier, respectively. $R_i(m)$ is modelled as the additive white gaussian noise (AWGN) with mean 0 and variance σ_z^2.

2.2 The broad learning system (BLS)

As shown in Figure 1, the BLS is used as an alternative to deep network structure. The input data is mapped to the mapping features and the enhancement features. The output layer connects the feature and enhancement layers. The data is transformed using a linear mapping function connecting the input weight matrices to obtain the set of mapped features. The $\boldsymbol{Z}_i$ is the mapped feature, which is obtained by linear mapping and activation function, i.e.

$$\boldsymbol{Z}_i = \varphi_i(\boldsymbol{X}\boldsymbol{V}_{ei} + \alpha_{ei}), i = 1, 2, \ldots, n \tag{3}$$

where $X \in R^{a\times b}$ denotes the input sample data for model training, where a is the sample number and b is the dimension. The φ_i is the activation function of the feature node, α_{ei} denotes the input layer to the mapping feature layer. The mapping feature group is denoted as

$$\boldsymbol{Z}^n = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3, \ldots, \boldsymbol{Z}_m] \tag{4}$$

The enhancement nodes are obtained from the mapped nodes by the nonlinear mapping and the activation function transformation. Y_j is the jth set of augmented nodes, which is

$$Y_j = \boldsymbol{\xi}_j\left(\boldsymbol{Z}^n V_{hj} + \alpha_{hj}\right), j = 1, 2, \ldots, n \tag{5}$$

where $\boldsymbol{\xi}_j$ is the activation function of the enhancement node; V_{hj} denotes the random connection weight matrix of the j-th group of mapped feature nodes to the layer of enhancement nodes; α_{hj} is the bias matrix of the j-th group, and the group of enhancement nodes obtained by the n-transform is denoted by

$$\mathbf{D}^n = [\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \ldots, \mathbf{D}_n] \tag{6}$$

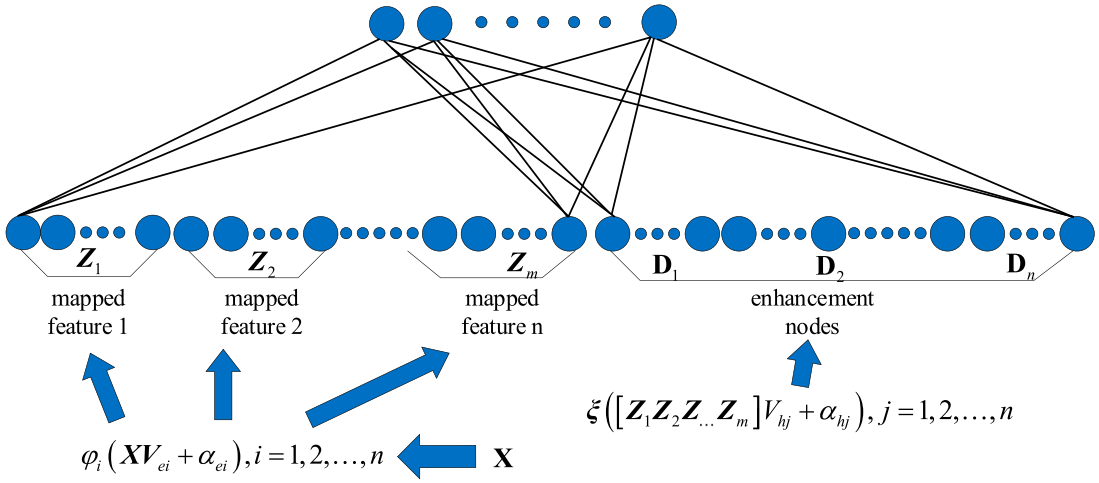


FIGURE 1
The typical network structure of BLS.

TABLE 1 Model parameter settings for the channel prediction method based on adaptive broad learning network.

Symbol	Meaning	Value	Symbol	Meaning	Value
f_d	Carrier frequency	2.4 GHz	N_{FFT}	Total number in FFT	64
BW	Bandwidth	2 MHz	T_{IDFT}	IDFT/DFT period	32us
MD	Modulation	QPSK	T_g	Guard interval length	9us
f_{OFDM}	OFDM symbol rate	30 kHz	N_t	Number of transmitting antennas	1
CR	Code rate	1/2	N_r	Number of receiving antennas	1
DR	Data rate	1.2Mbps	Nss	Quantity of data streams	1
K	Subcarrier number per OFDM symbol	50	N_{DATA}	DATA OFDM symbol count for each frame	1
N_{SPS}	Pilot subcarrier per OFDM symbol	4	f_r	Sample rate	2 MHz

Then, the outputs of the feature nodes and enhancement nodes are combined, i.e., $F = [Z_m \mid D_n]$. The weighted mapping of F, forms the output of the network $\hat{H} = FV$. The mapped feature nodes and enhancement nodes are used as inputs to the BLS and the feature vectors are solved. V denotes the weight matrix from the input layer to the output layer of the system, which is solved by solving the ridge regression, i.e.,

$$\mathbf{V} = \left(\lambda \boldsymbol{I} + \mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\boldsymbol{F}\boldsymbol{H}^{\mathrm{T}} \tag{7}$$

where the H denotes the regularization factor and I denotes the unit matrix.

3 Channel predictionbased of adaptive BLS

3.1 ABLN

The hidden layer of a typical BLN often has a large number of neurons. After using Eq. 5 to estimate the BLS network's output weight matrix, the estimated value is generally an ill-conditioned solution, which has a large amplitude. Therefore, the ill-conditioned solution problem of the output weight matrix is a problem that cannot be ignored in the training process of the BLN (Zhou, 2013). To improve the ABLN's generalization capabilities throughout the learning process, this study proposes using a hybrid generalization network to estimate the output weight matrix. The initial layer of the hybrid generalization network looks like this:

$$J_k = \arg\min_{\hat{Q}_{:,i}} \left\{ \left\| \mathbf{Y}_i - \mathbf{Z}\mathbf{Q}_{:,i} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,i} \right\|_2^2 + \lambda_{\frac{1}{2}} \left\| \mathbf{Q}_{:,i} \right\|_{\frac{1}{2}}^{\frac{1}{2}} \right\}_k \tag{8}$$

where $i = 1,2,3, ., I$, $\mathbf{Y}_i$ denotes the target output corresponding to the i-th step prediction, $\mathbf{Q}_{:,i}$ is the estimated value of the target output weight matrix corresponding to the i-th prediction. $\|^*\|_2$ and $\|^*\|_{\frac{1}{2}}$ denote the l_2 norm and $l_{1/2}$ norm, respectively, and $\lambda_{\frac{1}{2}}$ is the generalization coefficient of $l_{1/2}$ norm. Therefore, the first layer is further expressed as

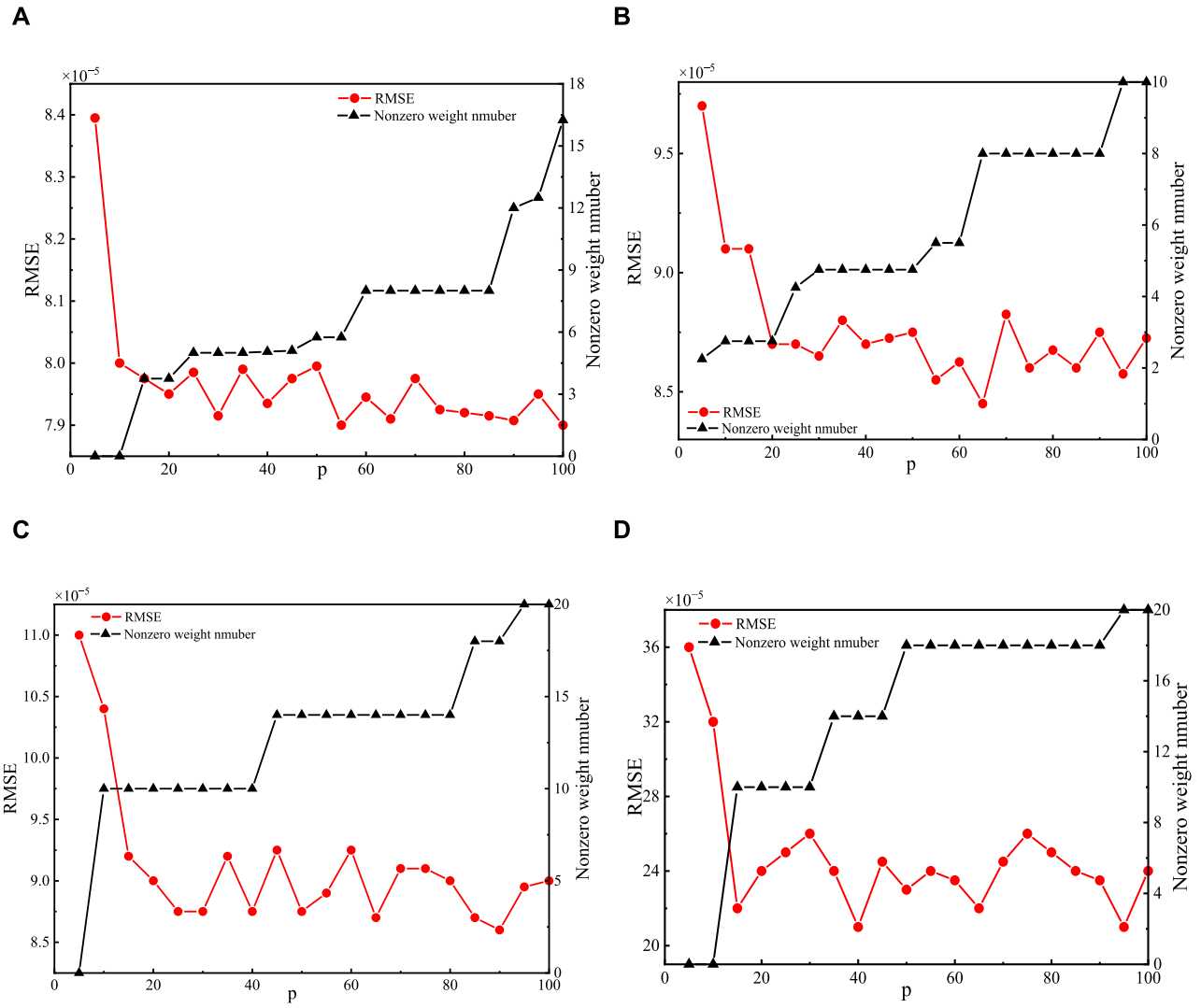


FIGURE 2
Non-zero weight curves and RMSE curves in ABLN with varying input dimensions: The actual and hypothetical components of the indoor scenario are shown in **(A,B)**, respectively; The actual and imagined components of the outdoor landscape are shown in **(C,D)**.

$$J_k = \arg\min_{\hat{Q}_{:,i}} \left\{ \left\| \mathbf{Y}_i - \mathbf{Z}\mathbf{Q}_{:,i} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,i} \right\|_2^2 + \lambda_{\frac{1}{2}} \left\| \mathbf{Q}_{:,i} \right\|_{\frac{1}{2}}^{\frac{1}{2}} \right\}_k$$
$$= \arg\min_{\hat{Q}_{:,i}} \begin{Bmatrix} \left\| \mathbf{Y}_1 - \mathbf{Z}\mathbf{Q}_{:,1} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,1} \right\|_2^2 + \lambda_{\frac{1}{2}} \left\| \mathbf{Q}_{:,1} \right\|_{\frac{1}{2}}^{\frac{1}{2}} + \\ \left\| \mathbf{Y}_2 - \mathbf{Z}\hat{\mathbf{Q}}_{:,2} \right\|_2^2 + \lambda_1 \left\| \hat{\mathbf{Q}}_{:,2} \right\|_2^2 + \lambda_{\frac{1}{2}} \left\| \hat{\mathbf{Q}}_{:,2} \right\|_{\frac{1}{2}}^{\frac{1}{2}} + \\ \vdots \\ \left\| \mathbf{Y}_I - \mathbf{Z}\mathbf{Q}_{:,I} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,I} \right\|_2^2 + \lambda_{\frac{1}{2}} \left\| \mathbf{Q}_{:,I} \right\|_{\frac{1}{2}}^{\frac{1}{2}} \end{Bmatrix} \tag{9}$$

Therefore, the above Eq. 9 is equivalent to solve the following issue, i.e.,

$$J_k = \arg\min_{\hat{Q}_{:,i}} \left\{ \left\| \mathbf{Y}_i - \mathbf{Z}\mathbf{Q}_{:,i} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,i} \right\|_2^2 + \lambda_{\frac{1}{2}} \left\| \mathbf{Q}_{:,i} \right\|_{\frac{1}{2}}^{\frac{1}{2}} \right\}_k \tag{10}$$

where $i = 1, 2, 3, ..., I$. Then, the above equation is rewritten as

$$J_k = \arg\min_{\hat{Q}_{:,i}} \left\{ \left\| \mathbf{Y}_i - \mathbf{Z}\hat{\mathbf{Q}}_{:,i} \right\|_2^2 + \lambda_1 \left\| \hat{\mathbf{Q}}_{:,i} \right\|_2^2 + \lambda_{\frac{1}{2}} \left\| \mathbf{Q}_{:,i} \right\|_{\frac{1}{2}}^{\frac{1}{2}} \right\}_k$$
$$= \arg\min_{\hat{Q}_{:,i}} \left\{ \left(\left\| \begin{bmatrix} \mathbf{Y}_i \\ 0 \end{bmatrix} - (1+\lambda_2)^{-\frac{1}{2}} \begin{pmatrix} \mathbf{Z} \\ \sqrt{\lambda_2} I \end{pmatrix} \hat{\mathbf{Q}}_{:,i} \right\|_2^2 + \frac{\lambda_2}{\sqrt{1+\lambda_{1/2}}} \left\| \mathbf{Q}_{:,i} \right\|_{\frac{1}{2}}^{\frac{1}{2}} \right) \right\}_k \tag{11}$$
$$= \arg\min_{\hat{Q}_{:,i}} \left\{ \left\| \hat{\mathbf{Y}}_i - \hat{\mathbf{Z}}\hat{\mathbf{Q}}_{:,i}' \right\| + \boldsymbol{\lambda}_b \left\| \mathbf{Q}_{:,i} \right\|_{\frac{1}{2}}^{\frac{1}{2}} \right\}_k$$

where

$$\hat{\mathbf{Y}}_i = \begin{bmatrix} \mathbf{Y}_i \\ 0 \end{bmatrix} \tag{12}$$

$$\hat{\mathbf{Z}} = (1+\lambda_2)^{-\frac{1}{2}} \begin{pmatrix} \mathbf{Z} \\ \sqrt{\lambda_2} I \end{pmatrix} \tag{13}$$

TABLE 2 One step prediction performance.

Scenario	Database	Model	Regularization parameter	RMSE	Average sparsity degree (%)	Average consuming time (s)
indoor scenario	Real component	BLN	-	9.54e−4	100	0.0258
		LBLN	3.9e−4	7.54e−4	3.05	1.5678
		RBLN	2.6e−4	4.87e−4	100	0.0258
		EBLN	2.8e−4,1.1e−4	6.71e−4	3.05	3.7364
		ABLN	1.4e−6,2.7e−6	1.65e−4	3.05	0.8612
	Imaginary component	BLN	-	5.11e-4	100	0.0252
		LBLN	5.6e−5	3.38e-4	3.13	1.4521
		RBLN	4.8e−5	3.38e-4	100	0.0198
		EBLN	0.9e−5,2.6e−6	2.41e-4	3.13	2.8974
		ABLN	8e−7,4e−6	1.64e−4	3.13	0.4874
Outdoor scenario	Real component	BLN	-	2.86e-4	100	0.0109
		LBLN	1.1e-4	1.58e-4	3.2	0.8124
		RBLN	5.3e−4	2.14e-4	100	0.0137
		EBLN	2.6e−6,1.1e−5	1.67e-4	2.05	2.9485
		ABLN	1.3e−6,2e−6	8.13e−5	1.99	0.2712
	Imaginary component	BLN	-	3.69e-4	100	0.0295
		LBLN	2.2e−5	1.63e-4	3.2	1.1566
		RBLN	4.8e−5	1.88e-4	100	0.0269
		EBLN	1.1e−5,2.8e−6	1.21e-4	2.16	3.8573
		ABLN	0.9e−6,1.8e−6	1.27e−4	2.11	0.2433

$$\boldsymbol{\lambda}_b = \frac{\lambda_2}{\sqrt{1+\lambda_{1/2}}} \tag{14}$$

Equation 11 is an optimization problem about $l_{1/2}$ norm. However, the $l_{1/2}$ norm is the non-convex and non-smooth, and some common optimization solution methods, i.e., the Newton's method and simulated Newton's method are difficult to solve the optimization Eq. 11. The coordinate descent method is utilized to solve Eq. 11 and its computation process is shown in Algorithm 2. Then the adaptive weighting factor $\delta_{:,i}$ for the first layer is

$$\delta_{:,i} = \left[\left| \mathbf{Q}_{:,i} \right| + \frac{1}{(M+L)} \right]^{-\tau} \tag{15}$$

where $\tau \in \mathbf{N}_+$ is the adaptive adjustment factor. The second layer of the output weight matrix estimation network for the BLN is:

$$J_k = \arg\min_{Q_{:,i}} \left\{ \sum_{i=1}^{I} \left(\begin{array}{l} \left\| \mathbf{Y}_i - \mathbf{Z}\mathbf{Q}_{:,i} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,i} \right\|_2^2 + \\ \lambda_2 \sum_{j=1}^{M+L} \left(\delta_{j,i}\, \mathbf{Q}_{j,i} \right) \end{array} \right) \right\}_k \tag{16}$$

where $\mathbf{Q}_{:,i}$ is the output weight matrix corresponding to the i-th prediction, $\mathbf{Q}_{j,i}$ denotes the jth element of the output weight matrix corresponding to the i-th step prediction, and λ_1 and λ_2 are the non-zero penalty coefficients, respectively. Eq. 16 can be further rewritten as

$$\begin{aligned} J_k &= \arg\min_{Q_{:,i}} \left\{ \sum_{i=1}^{I} \left(\begin{array}{l} \left\| \mathbf{Y}_i - \mathbf{Z}\mathbf{Q}_{:,i} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,i} \right\|_2^2 \\ +\lambda_2 \sum_{j=1}^{M+L} \left(\delta_{j,i}\, \mathbf{Q}_{j,i} \right) \end{array} \right) \right\}_k \\ &= \arg\min_{Q_{:,i}} \left\{ \begin{array}{l} \left\| \mathbf{Y}_1 - \mathbf{Z}\mathbf{Q}_{:,1} \right\|_2^2 + \lambda_2 \left\| \mathbf{Q}_{:,1} \right\|_2^2 + \lambda_1 \sum_{j=1}^{M+L} \left(\delta_{j,1}\, \mathbf{Q}_{j,1} \right) \\ + \left\| \mathbf{Y}_2 - \mathbf{Z}\mathbf{Q}_{:,2} \right\|_2^2 + \lambda_2 \left\| \mathbf{Q}_{:,2} \right\|_2^2 + \lambda_1 \sum_{j=1}^{M+L} \left(\delta_{j,2}\, \mathbf{Q}_{j,2} \right) \\ \vdots \\ \left\| \mathbf{Y}_I - \mathbf{Z}\mathbf{Q}_{:,I} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,I} \right\|_2^2 + \lambda_2 \sum_{j=1}^{M+L} \left(\delta_{j,i}\, \mathbf{Q}_{j,I} \right) \end{array} \right\} \end{aligned} \tag{17}$$

It can be shown that the loss function (17) of the second layer output weight matrix estimation network can also be converted into an I optimization problem with respect to the l2 norm and the adaptive weighting factors, i.e.,

$$J_k = \arg\min_{Q_{:,i}} \left\{ \left\| \mathbf{Y}_i - \mathbf{Z}\mathbf{Q}_{:,i} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,i} \right\|_2^2 + \lambda_2 \sum_{j=1}^{M+L} \left(\delta_{j,i}\, \mathbf{Q}_{j,i} \right) \right\}_k \tag{18}$$

TABLE 3 Multistep prediction performance.

Scenario	Database	Model	Average sparsity degree (%)	Average consuming time (s)
indoor scenario	Real component	BLN	100	0.0253
		LBLN[33]	20.8	2.1187
		RBLN	100	0.0162
		EBLN	15.63	1.7548
		ABLN	15.09	1.2869
	Imaginary component	BLN	100	0.0169
		LBLN[33]	17.69	1.4467
		RBLN	100	0.0089
		EBLN	16.64	3.3716
		ABLN	15.37	1.2285
Outdoor scenario	Real component	BLN	100	0.0243
		LBLN[33]	5.28	0.9624
		RBLN	100	0.0164
		EBLN	3.68	3.2686
		ABLN	4.47	0.4538
	Imaginary component	BLN	100	0.0245
		LBLN[33]	4.59	1.2572
		RBLN	100	0.0142
		EBLN	4.46	3.1453
		ABLN	4.29	0.6941

where $i = 1, 2, 3, \ldots, I$. Eq. 17 can be further derived as:

$$J_k = \arg\min_{Q_{:,i}} \left\{ \left\| \mathbf{Y}_i - \mathbf{Z}\mathbf{Q}_{:,i} \right\|_2^2 + \lambda_1 \left\| \mathbf{Q}_{:,i} \right\|_2^2 + \lambda_2 \sum_{j=1}^{M+L} \left(\delta_{j,i} \mathbf{Q}_{j,i} \right) \right\}_k$$

$$= \arg\min_{Q_{:,i}} \left\{ \left(\left\| \begin{bmatrix} \mathbf{Y}_i \\ 0 \end{bmatrix} - (1+\lambda_2)^{-\frac{1}{2}} \begin{pmatrix} \mathbf{Z} \\ \sqrt{\lambda_2} I \end{pmatrix} \mathbf{Q}_{:,i} \right\|_2^2 + \frac{\lambda_2}{\sqrt{1+\lambda_2}} \sum_{j=1}^{M+L} \left(\delta_{j,i} \mathbf{Q}_{j,i} \right) \right) \right\}_k \tag{19}$$

$$= \arg\min_{Q_{:,i}} \left\{ \left\| \mathbf{Y}_i' - \mathbf{Z}' \mathbf{Q}_{:,i}' \right\| + \boldsymbol{\lambda}_s \sum_{j=1}^{M+L} \left(\delta_{j,i} \mathbf{Q}_{j,i} \right) \right\}_k$$

Where $\mathbf{Y}_i'$, $\mathbf{Z}'$, $\mathbf{Q}_{:,i}'$ and $\boldsymbol{\lambda}_s$ are defined as follows:

$$\mathbf{Y}_i' = \begin{bmatrix} Y_j & \mathbf{0} \end{bmatrix}^{\mathrm{T}} \tag{20}$$

$$\mathbf{Z}' = (1+\lambda_2)^{-\frac{1}{2}} \begin{bmatrix} z & \sqrt{\lambda_2} I \end{bmatrix}^T \tag{21}$$

$$\mathbf{Q}_{:,i}' = \sqrt{1+\boldsymbol{\lambda}_2} w_i \tag{22}$$

$$\lambda_s = \frac{\lambda_1}{\sqrt{1+\lambda_2}} \tag{23}$$

where I is the unit matrix. It can be shown that Eq. 18 is an optimization problem with respect to the adaptive weighting factors. Further definition.

$$\mathbf{Q}_{j,i}'' = \delta_{j,i} \mathbf{Q}_{j,i}' \tag{24}$$

$$\mathbf{Z}'' = \frac{\mathbf{Z}'}{\delta_{:,i}} \tag{25}$$

Eq. 19 can be further rewritten as:

$$J_k = \arg\min_{Q_{:,i}'} \left\{ \left\| \mathbf{Y}_i' - \mathbf{Z}' \mathbf{Q}_{:}'' \right\|_2^2 + \boldsymbol{\lambda} \sum_{j=1}^{M+L} \left(\mathbf{Q}_{j:}'' \right)_i \right\}_k$$
$$= \arg\min_{Q_{:,i}'} \left\{ \left\| \mathbf{Y}_i' - \mathbf{Z}' \mathbf{Q}_{:}'' \right\|_2^2 + \boldsymbol{\lambda} \left\| \mathbf{Q}_{:}'' \right\| \right\}_k \tag{26}$$

Eq. 26 is an optimization problem with respect to the L_1-paradigm. Currently, some methods are available, such as the Newton's method and the gradient descent method. All of the approaches, though, need the solution variables' derivatives. In this study, we answer Eq. 27 using the minimal angle regression (MAR) approach, which avoids the derivatives of the variables. When $\mathbf{Q}_{:,i}''$ is obtained, the output weight matrix $\mathbf{Q}_{:,i}$ corresponding to the ith step prediction is

$$\mathbf{Q}_{:,i} = \frac{\mathbf{Q}_{:,i}''}{\left(\delta_{:,i} \sqrt{1+\lambda_2} \right)} \tag{27}$$

Equation 27 is further modified as (Rodan, 2012):

$$\mathbf{Q}_{:,i} = \frac{\sqrt{1+\lambda_2}}{\delta_{:,i}} \mathbf{Q}_{:,i}'' \tag{28}$$

Therefore, the output weight matrix $\mathbf{Q}_{out}$ is

Input: training output matrix $\mathbf{Y}$, training input matrix $\mathbf{Z}$, penalty coefficients λ_1, λ_2 and $\lambda_{1/2}$, adaptive factors τ, maximum prediction step number I
Output: weight matrix $\mathbf{Q}_{out}$
Step 1: For $i = 1, 2, 3, \ldots, I$
Step 2: Solve Eq. 8 to obtain $\mathbf{Q}_{:,i}$;
Step 3: Calculate the $\delta_{j,i}$ by Eq. 15;
Step 4: Calculate the modified $\mathbf{Y}_i^{'}$ by Eq. 20;
Step 5: Calculate the modified $\mathbf{Z}''$ by Eq. 21;
Step 6: Solve Eq. 26 by the minimum angle regression method;
Step 7: Estimate by Eq. 28.
end
Step 8: Construct the output weight matrix $\mathbf{Q}_{out}$ by Eq. 29
Step 9: Output $\mathbf{Q}_{out}$;

Algorithm 1. Calculation process of the ABLN.

Inputs: input: training output matrix $\mathbf{Y}$, training output matrix $\mathbf{Z}$, non-zero penalty coefficients $\lambda_{1/2}$, empty matrix $\mathbf{Q}_{:,i}^{0}$
Output: $\mathbf{Q}_{:,i}$;
Step 1: $\mathbf{Q}_{:,i} = \mathbf{Q}_{:,i}^{0}$;
Step 2: For $j = 1, 2, 3, \ldots, L + M$;
Step 3: $\mathbf{Q}_{:,i}^{0} = \mathbf{Q}_{:,i}$;
Step 4: Estimate $\mathbf{Q}_{:,i}$, $p = 1, 2, 3, \ldots, j$;
Step 5: Judge $\sum_{j=1}^{M+L} |\mathbf{Q}_{j,i}^{0} - \mathbf{Q}_{j,i}| \leq \varepsilon$. If no, then jump back to step 3; If yes, then jump to step 2.
End

Algorithm 2. Coordinate descent computation procedure.

$$\mathbf{Q}_{out} = [\mathbf{Q}_1, \mathbf{Q}_2, \cdots, \mathbf{Q}_I] \tag{29}$$

The computational pseudo-code for the ABLN is shown in Algorithm 1 and the computational pseudo-code for the coordinate descent method is shown in Algorithm 2.

4 Simulation and discussion

These related parameters are part of the channel prediction approach for the DAMS based on ABLN that is looked at in this paper; they are shown in Table 1. Both the interior and outdoor scenarios are taken into consideration in order to assess how well ABLN performs.

1) The indoor scenario: Among the pertinent features are the maximum Doppler shift of 80 Hz and the total of six transmission routes; the delay and power are, respectively, (0, 120, 430, 860, 1030, 1470) ns and (0, −1.6, −0.8, −9, −9.4, −2.8) dB.
2) The outdoor scenario: The relevant parameters for the Nakagami-m channel scenario are as follows: The maximum Doppler shift is 20 Hz, and m is set to 5. Then, there are a total of four transmission lines, each having the following powers and de-lays: 0 dB and (0, 130, 230, 1180) ns, respectively.

This work also examines several of the current BLN-based methodologies, such as the fundamental BLN (Zhao and Lu, 2023), the lasso regularized BLN (LBLN) (Duan and Xu, 2022), the BLN with the ridge regularization (RBLN) (Wang, 2022) and the elastic network BLN (EBLN) (Ding and Xie, 2023). The following three performance metrics are considered in this paper.

1) The difference between the ideal and expected CSI is represented by the Root Mean Square Error, or RMSE. Thus, the prediction performance is evaluated in this paper using RMSE. Better predictive performance of the model is indicated by a reduced root mean square error.
2) The examined model's ability to generate sparse output weight matrices is shown by its average sparsity. Again, a sparse output weight matrix demonstrates the significant improvement in the model's memory use.
3) The average time spent is used to determine how complex the computation is during the output weight matrix estimation procedure. When the average time to estimate the output weight matrix is lower for a given model, we can conclude that the model has less computational complexity.

These three metrics are computed and examined in terms of one step and multi-step prediction in this section. The three measures are averaged over ten runs to remove randomness. The average sparsity for one step prediction is the mean number of non-zero elements in the projected output weight matrix after 10 iterations. The average sparsity in multi-step prediction is defined as the average rate of non-zero elements in the predicted output weight matrix from one step prediction to h-step prediction. Besides, in the simulation, the relevant parameter parameters of the ABLN include the neuron number of the feature layer 100, the neuron number 50, the input scaling factor 0.01 of the feature layer. Random generation is used to create both the input weight matrix and the internal weight matrix within the interval [-1, 1]. Each OFDM symbol has 60 subcarriers, as seen in Table 1, and 120 predictors might theoretically be used to complete the IDFT process. To keep things simple, the model utilized in this research is only evaluated using the CSI samples of the first subcarrier. The model is trained in ABLN using 5000 frames, and its performance is then assessed using the 1000 frames that follow.

4.1 Input dimensions of ABLN

The model's input dimensions have an impact on the prediction performance. As a result, in two channel cases, the input dimensions of ABLN must be estimated. We present simulation results and a discussion of input dimensions in ABLN in this sub-section. Table 1 provides these regularization parameters. The x-axis of Figure 2, which shows the RMSE curves and non-zero weight curves of the predicted output weight matrices of both evaluation models in the two scenarios given, represents the input dimension of the ABLN, or the total CSI samples utilized to forecast the next CSI sample. The relationship between the RMSE and the input dimension p is thus shown by the red curve with a solid red circle, and the RMSE is represented by the left vertical axis. Similarly, the black curve labeled with black solid triangles represents the relationship between the

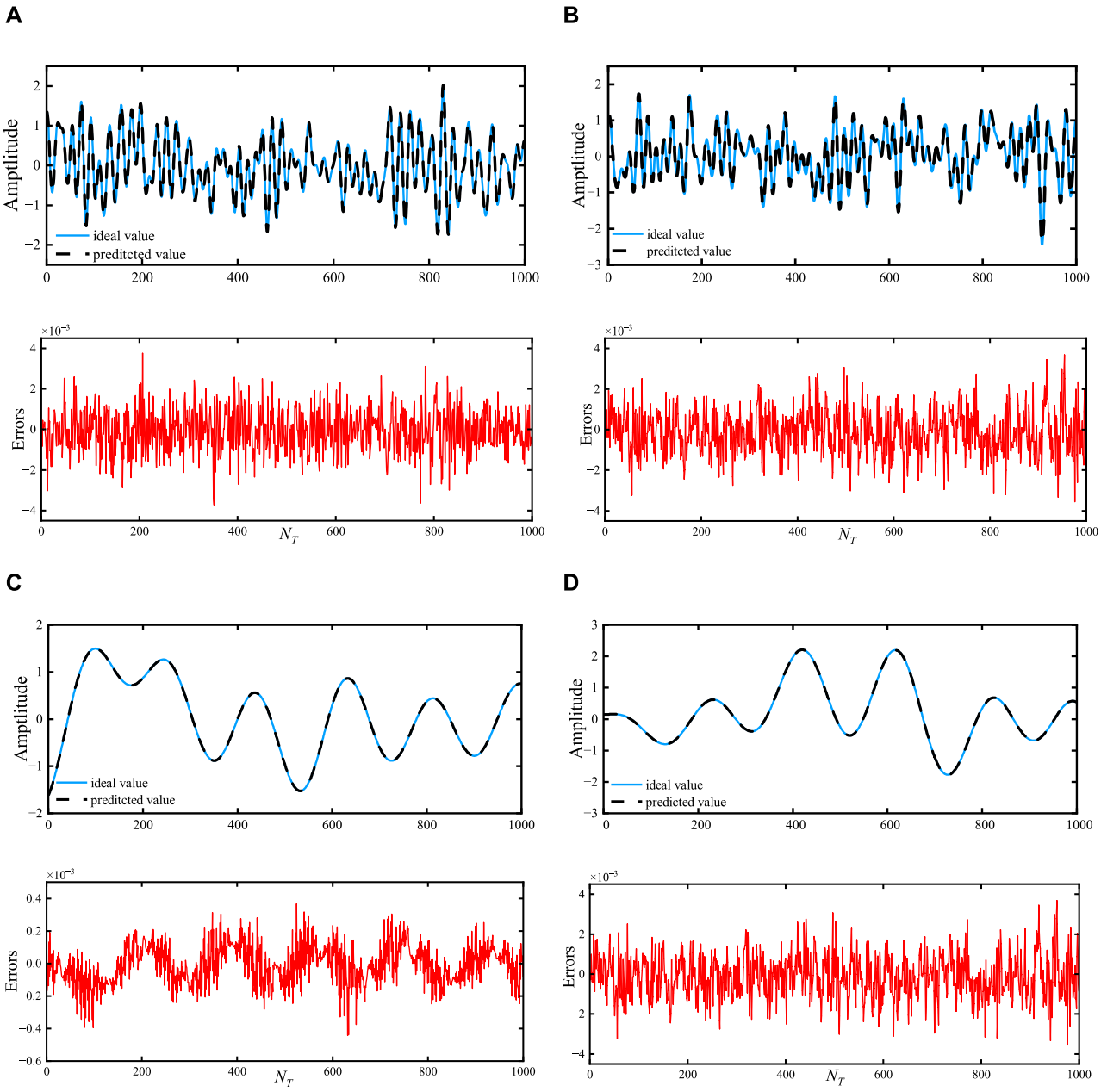


FIGURE 3
The predicted real and imaginary components of the inside scene are represented by the scenes **(A,B)**, respectively, and the predicted real and imaginary elements of the outside scene by the scenes **(C,D)**, respectively.

number of non-zero components and the input dimension p, and the number of non-zero elements in the expected output weight matrix is displayed on the right vertical axis.

Generally, as the input dimension of the ABLN increases, the RMSE tends to decrease. Furthermore, when the input dimension increases, the projected output weight matrix's non-zero members also grow. In Figure 2A, due to the increasing input dimensions, the RMSE curve of the real part of the CSI samples tends to stabilize in the indoor scenario, and the final RMSE is about 7.9e-4. Because of the increasing input dimensions, the redundancy information in Figure 2A raises the number of non-zero components of the projected output weight matrix. As a result, the lowering.

RMSE has a significant impact on the predicted output weight matrix's sparsity. The primary ex-planation behind this is that as input dimensions grow, the ABLN model has access to more data and may perform better when making predictions. The number of non-zero items in the predicted output weight matrix rises as a result of the increased redundant information brought about by the larger input dimensions. Generally, the actual fraction of the CSI samples in the interior situation should have an input dimension of about 35. Subsequently, regarding the fictitious portion of the CSI samples in the indoor setting, the stabilized RMSE is about 8.8e-5, and the appropriate input dimension is about 35. Thus, in the indoor scenario, we set the input dimensions to 35 for both the imaginary

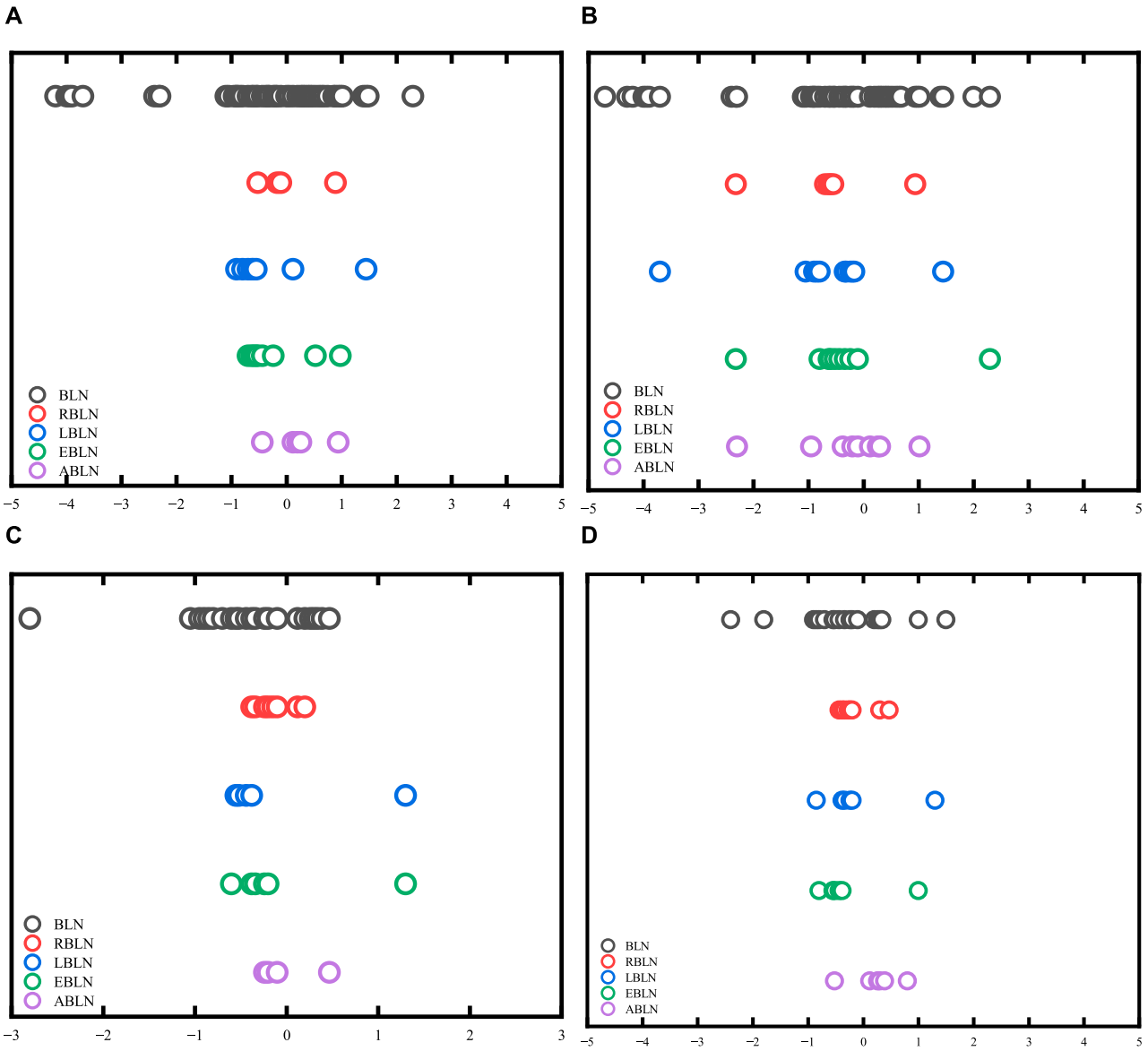


FIGURE 4
The estimated output weight matrix's distribution range. in the indoor scene, **(A,B)** are the real and imaginary parts, respectively; **(C,D)** the real and imaginary parts in the outdoor scene.

portion of the CSI samples and the real values. Then the external situation Figure 2 display the non-zero weight curves and RMSE curves. The steady RMSEs of the real and imaginary sections of the CSI samples are, as can be shown from Figures 2C, D, approximately 9.2e-5 and 2.4e-4, respectively, just like in the indoor scene. Therefore, we also adjust the input dimensions of the real and imaginary sections of the CSI samples in the outdoor scene set to 35 to reduce the amount of non-zero weights in the estimated weight matrices.

4.2 One step prediction

Table 2 provides the regularization settings in ABLN for one step prediction. Moreover, the cost parameter c in the c-SVC, gamma in the kernel function, epsilon p in the SVM loss function, and 25, 0.001 and 2e-5 for the outside scenario are assigned the values 25, 0.006, and 2e-5. c, g and p for these evaluated model predictive performances are shown in Table 3. Figure 3 shows the ideal CSI curves, projected CSI curves, and error curves of the ABLN in the indoor scene and outdoor scenario for the real and imaginary halves of the CSI sample. The distribution range of these assessment models' estimated output weight matrices is displayed in Figure 4.

The predicted CSI in ABLN, as illustrated in Figure 3A, closely resembles the ideal CSI with a maximum error of 2e-3, whereas Figure 3B shows a maximum error of 3e-3 for the imaginary component of the indoor scene. The correlation curves for the out-door scenario are displayed in Figures 3C, D. The predicted CSI curves, with a maximum error of 3e-4 in Figure 3C and 5e-3 for the imaginary component in Figure 3D, are observed to be satisfactorily accurate in matching the ideal curves. From

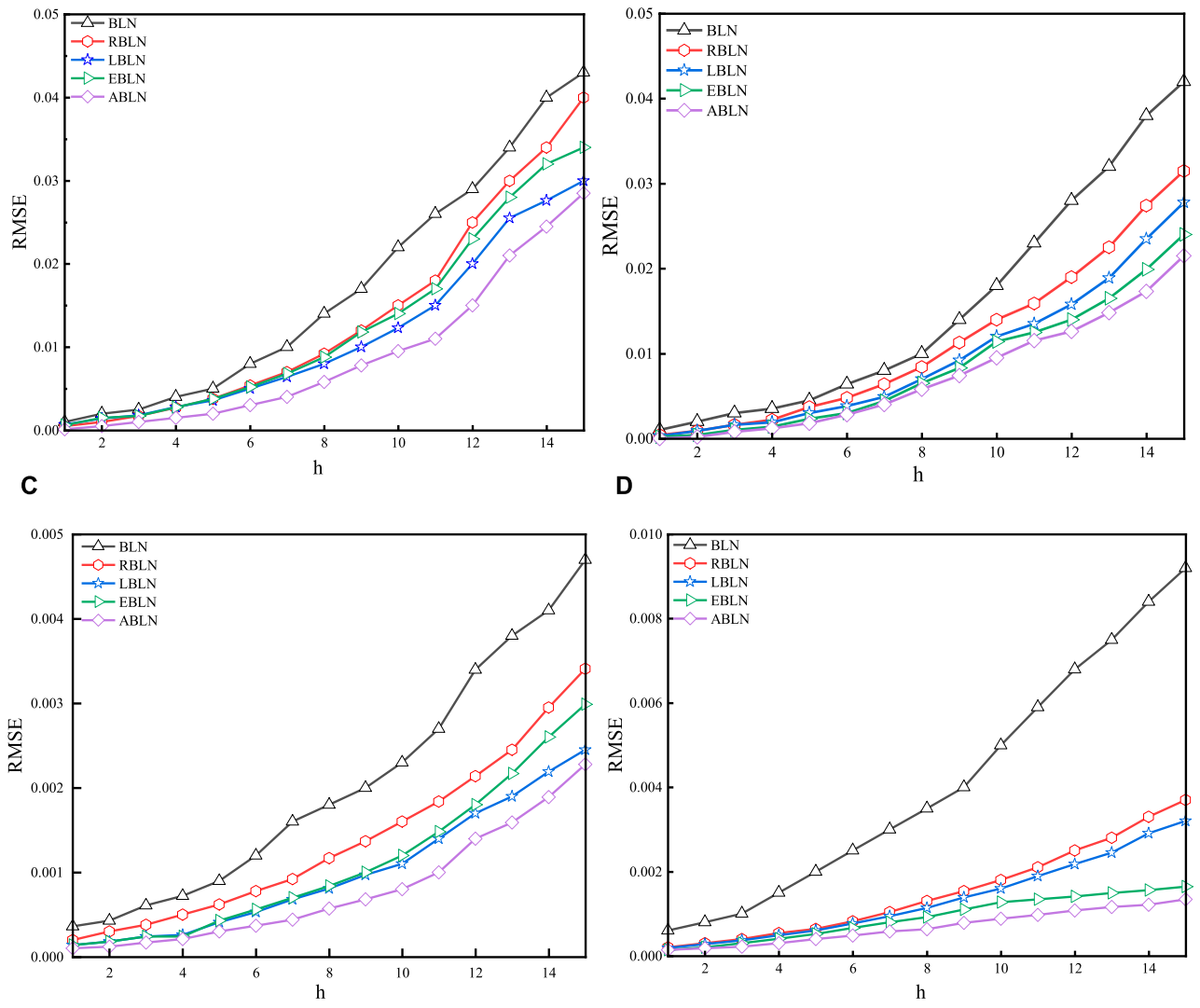


FIGURE 5
Multi-step prediction performance in indoor and outdoor scenarios: in the indoor scenario, **(A,B)** are real and imaginary parts, respectively; **(C,D)** real and imaginary parts in the outdoor scenario.

Figure 3, it can be seen that ABLN has relatively small prediction errors in one step prediction of indoor scenes *versus* outdoor scenes compared to other models. Thus, for the CSI samples in the indoor scene and the outdoor scene, we can obtain excellent one step ABLN prediction performance. The one step prediction performance is shown in Table 2. Table 2 also includes the regularization parameters of the BLN-based models for the two scenarios. The assessed models' average sparsity for the actual portion of the CSI samples in the in-door scenario is 3.05%, except for BLN and RBLN. But ABLN has the lowest RMSE (1.65e-4) compared to LESN (7.54e-4) and RESN (4.87e-3) and EESN (6.71e-4). It is calculated that, in terms of one step prediction performance, ABLN has a maximum improvement of about 96.49% with the other evaluated models. Furthermore, the amount of time needed to compute the output weight estimate matrix is also less in ABLN, with an average consumption time of about 0.8612 s, compared to 1.5678 s in LBLN and 3.7364 s in EBLN. Specifically, there is a considerable reduction in the computational time needed for the adaptive elasticity network in ABLN to estimate the input weight matrix. Although BLN and RBLN need less computing time (i.e., 0.0258), they do not yield a sparse output weight matrix. In the indoor situation, ABLN exhibits the best RMSE of 1.64e-4 and the same average sparsity of 3.13% for the imaginary part of the CSI samples. It also requires relatively little processing time, 0.4874 s. It is calculated that, in terms of the one step prediction performance, the maximum improvement of ABLN with respect to the other evaluated models is about 92.96%. ABLN requires smaller regularization parameters compared to other evaluation models. As a result, ABLN creates good sparse output weight matrices, has strong one step pre-diction performance, and uses less computing time to estimate the output weight matrices for CSI data in indoor environments. In the outdoor scenario, BLN exhibits the lowest performance for the genuine half of the CSI samples and requires a large amount of time to train the entire

model; in contrast, ABLN requires very little computational time (0.2712 s) and has an ideal RMSE of 8.13e-5 and 1.99% mean sparsity. ABLN also has the same mean sparsity (e.g., 2.11%) for the imaginary component with an ideal RMSE (i.e., 1.27e-4), and a very short computation time (i.e., 0.2433 s). The highest improvement of the ABLN for the real and imaginary sections of these CSI samples compared to the other examined models in outdoor scenarios, is around 95.01% and 97.31% in terms of one step prediction performance. Therefore, ABLN can offer good one step prediction

Performance and shorter computing time for estimating the output weight matrices in two given communication scenarios for similar sparse output weight matrices.

Figure 4 displays the range of the estimated output weight matrix's distribution in the BLN-based assessment model. The projected output weight matrix in BLN has a substantial size, as seen in Figure 4A, and it roughly ranges from −5 to 5. Thus indicating that the generalization ability of the input CSI samples is unsatisfactory. In addition, the output weight matrix of the BLN is not sparse. The output weight matrix of the RBLN is also not sparse, despite having an estimated output weight matrix with the smallest range of magnitude. Furthermore, even though the LBLN's output weight matrix is sparse, its magnitude range exceeds the magnitude range of the RBLN. Com-pared with the output weight matrix of EBLN, the output weighting matrix typically approaches zero in ABLN. In terms of sparsity and magnitude range of the output weight matrix, ABLN has a clear advantage, which also indicates that ABLN has good generalization ability to the input CSI samples.

4.3 Multistep prediction

These parameters are the same as those in Table 3 for multi-step forecasts. Figure 5 displays the correlation curves for these assessed models' 1-step to 15-step predictions for the two provided scenarios. Overall, the ABLN has the best multi-step prediction performance better compared to BLN, RBLN, LBLN and EBLN for the two given scenarios. In the indoor scenario, for the real and imaginary parts of the CSI samples, the improvement ratios of ABLN with the above evaluation models in the 15th step are about 80.14% and 82.03%, respectively, while in the outdoor scenario, these improvement ratios are about 90.87% and 97.42%, respectively. Moreover, ABLN has the best multi-step prediction performance for the BLN-based assessment model, particularly for the real portion of the CSI samples (Figures 5A, C). For the imaginary part of the two given scenes (as shown in Figures 5B, D), among the evaluation models, ABLN performs better in multi-step prediction. In the indoor scene, the 15th step prediction improvement rates of BLN-based ABLN compared to other evaluated models are approximately 35.28% and 28.94% for the real and imaginary parts of the CSI sample, respectively. Conversely, in the outdoor scene, these improvement rates are approximately 48.35% and 49.46%, respectively. Thus, as far as prediction performance is concerned, the ABLN has satisfactory multi-step performance for both real and imaginary parts in both given scenarios. The results for the evaluated models in terms of average sparsity and average time consumption are presented in Table 3. In this paragraph, the average sparsity and average consumption time are defined using the mean of predictions made across stages one through fifteen. In the indoor scenario, the output weight matrix in ABLN exhibits the optimal sparsity of 14.97% for the real part of the CSI samples, whereas the output weight matrices in BLN and BLN exhibit no sparsity. The BLN beats the other evaluation models, including the ABLN, in terms of the average time needed to estimate the output weight matrix. But as was already established, the BLN's output weight matrix is not sparse. The conclusions are like those of the BLN. Then, the ABLN is also better in estimating the average time consumed for the output weight matrix compared to LBLN and EBLN. Therefore, the ABLN exhibits excellent multi-step prediction performance for the authentic portion of CSI data in the indoor scenario, considering both the average time required for calculating the output weight array and the sparsity of the estimated output weight matrix. The corresponding outcomes in the outdoor scenario are comparable to those obtained for the actual portion of CSI samples in the in-door scenario.

5 Conclusion

The channel prediction problem for distribution automation masters is the main topic of this study. Specifically, we propose in this study an ABLN for predicting the CSI of distribution automation master communications in OFDM systems. In this study, we simulate and analyze communication scenarios in both indoor and outdoor environments. The indoor scenario considers the Nakagami channel, while the outdoor scenario uses the Rayleigh channel. Moreover, in the simulation part, one step, multi-step and local predictability proofs for CSI samples are implemented. After the simulation, we draw several inferences. First, the CSI samples inferred by least squares (LS) in orthogonal frequency division multiplexing (OFDM) symbol subcarriers have significant local predictability, especially under the condition of finite maximum Doppler shift. In both cases, the local predictability of the best CSI samples exceeds 94.95%. Furthermore, in both cases, ABLN exhibits excellent one step and multistep prediction performance for both the real and imaginary parts of the CSI samples. In one step prediction, ABLN provides well one step prediction performance and shorter computation time to estimate the output weight matrix. In multistep prediction, ABLN provides satisfactory multistep performance in both real and imaginary parts of two given scenarios. It is characterised by sparse output weight matrices with small magnitude and less computational time required to estimate the output weight matrix, especially in one step prediction. The maximum improvement in the prediction performance of ABLN is calculated to be about 96.49% compared to other models. In addition, the regularization factor of ABLN is smaller than the regularization parameters of other evaluated models. Therefore, ABLN has advantages in channel prediction based on adaptive distribution automation master communication and can be used for future adaptive wireless communication in distribution automation master. The channel prediction method proposed in this paper can provide highly accurate channel prediction information, which provides a guarantee to reduce the impact of outdated channel state information on wireless communication systems such as distribution network automation master stations. The optimization and modification of the ABLN further to improve the prediction performances of the system model are the works in the future.

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

PL: Formal Analysis, Writing–original draft, Writing–review and editing. RZ: Investigation, Software, Writing–review and editing. HF: Data curation, Methodology, Writing–review and editing. HW: Resources, Validation, Visualization, Writing–review and editing. ZY: Data curation, Methodology, Supervision, 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 research was funded by National Key Research and Development Program: Joint research and development and demonstration of collaborative energy management and operation optimization technology for "Belt and Road" National Urban Smart Energy Network (524608200157).

Conflict of interest

Authors PL, HF, and HW were employed by NARI Group Co., Ltd. Authors PL, HF, and HW were employed by NARI Technology Nanjing Control Systems Co., Ltd. Authors RZ and ZY were employed by Guangdong Power Grid Corporation.

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

Bai, Q., Wang, J., Zhang, Y., and Song, J. (2019). Deep learning based channel estimation algorithm over time selective fading channels. *IEEE T Cogn. Commun.* 6, 125–134. doi:10.1109/TCCN.2019.2943455

Chen, C. L. P., and Liu, Z. (2018). Broad learning system: an effective and efficient incremental learning system without the need for deep architecture. *IEEE T Neur. Net. Lear.* 29, 10–24. doi:10.1109/TNNLS.2017.2716952

Chen, C. L. P., Liu, Z., and Feng, S. (2019). Universal approximation capability of broad learning system and its structural variations. *IEEE T Neur. Net. Lear.* 30, 1191–1204. doi:10.1109/TNNLS.2018.2866622

Ding, H. F., Xie, Y. F., Xie, S. W., and Wang, J. (2023). A width learning system based on feature layer dense connectivity and attention mechanism and its application to zinc flotation process. *Cont. Appli.* 40, 111–120. doi:10.7641/CTA.2022.10753

Ding, T., and Hirose, A. (2013). "Fading channel prediction based on complex-valued neural networks in frequency domain," in IEEE International Symposium on Electromagnetic Theory, Hiroshima, Japan, May 2013, 640–643.

Ding, T., and Hirose, A. (2014). Fading Channel prediction based on combination of complex-valued neural networks and chirp Z-transform. *IEEE T Neur. Net. Lear.* 25, 1686–1695. doi:10.1109/TNNLS.2014.2306420

Duan, J. W., Xu, L. C., Quan, Y. J., Chen, L., and Chen, J. L. (2022). Bayesian width learning system based on graph regularization. *J. Intelli. Sci. Tech.* 4, 109–117. doi:10.11959/j.issn.2096–6652.202203

Feng, S., and Chen, C. L. P. (2018). Fuzzy broad learning system: a novel neuro-fuzzy model for regression and classification. *IEEE T Cy* 99, 414–424. doi:10.1109/TCYB.2018.2857815

Gu, Y. (2017). Review of development status and trend of distribution automation in China. *J. Electr. Engin* 5, 270–277. doi:10.12677/jee.2017.54033

Jiang, W., and Schotten, H. D. (2019). "Recurrent neural network-based frequency-domain Channel Prediction for wideband communications," in IEEE Vehicular Technology Conference, Kuala Lumpur, Malaysia, April 2019, 1–9. doi:10.1109/VTCSpring.2019.8746352

Jin, J. W., and Chen, C. L. P. (2018). Regularized robust Broad Learning System for uncertain data modeling. *Neurocomputing* 322, 58–69. doi:10.1016/j.neucom.2018.09.028

Kapoor, D. S., and Kohli, A. K. (2018). Channel estimation and long-range prediction of fast fading channels for adaptive OFDM system. *Int. J. Electron. Theor. Exp.* 105, 1451–1466. doi:10.1080/00207217.2018.1460871

Kurtz, R. B. S. A. (1987). Johan Håstad. Computational limitations of small-depth circuits. ACM doctoral dissertation awards. The MIT Press, Cambridge, Mass., and London, 1987, xiii + 84 pp. *J. Symbolic Log.* 53, 1259–1260. doi:10.2307/2274626

Lee, J., Park, C., and Roh, H. (2021). "Revisiting bluetooth adaptive frequency hopping prediction with a ubertooth," in international conference on information networking (ICOIN), Jeju Island, Korea (South), January 2021, 715–717. doi:10.1109/ICOIN50884.2021.9333996

Liang, J. Y. (2023). Research on decentralized breadth learning system with high communication efficiency. *Guilin Uni. Tech.* doi:10.27050/d.cnki.gglgc.2023.001071

Min, H., Shoubo, F., Philip, C. C. L., Meiling, X., and Tie, Q. (2018). Structured Manifold broad learning system: a Manifold perspective for large-scale chaotic time series analysis and prediction. *IEEE T Knowl. Data En.* 31, 1809–1821. doi:10.1109/TKDE.2018.2866149

Muhammad, S. J. S., Huseyin, A. H., TurkmenHaji, M. F., and Arslan, H. (2021). Wireless communication, sensing, and REM: a security perspective. *IEEE Open J. Commun. Soc.* 2, 287–321. doi:10.1109/OJCOMS.2021.3054066

Niu, G. Q., Yan, Y., Li, Y. C., Yao, X. J., Wang, C. M., and Gao, X. (2014). "Adaptive autoregressive prediction method for deep-space channel using kalman filter," in 2014 IEEE International Conference on Instrumentation and Measurement, Computer, Communication and Control, Harbin, China, September 2014, 533–538.

Pan, L., Han, Z., Shanshan, Z., and Feng, W. (2023). An optimal allocation method for power distribution network partitions based on improved spectral clustering algorithm. *Int. J. Intelli. Real-Time Automation* 123, 106497–106514. doi:10.1016/j.engappai.2023.106497

Pao, Y. H., and Takefuji, Y. (2021). Functional-link net computing: theory, system architecture, and functionalities. *Computer* 25, 76–79. doi:10.1109/2.144401

Pu, X., and Li, C. (2021). Online semi-supervised broad learning system for industrial fault diagnosis. *IEEE T Ind. Inf.* 17, 6644–6654. doi:10.1109/TII.2020.3048990

Rodan, A., and Tino, P. (2012). Simple deterministically constructed cycle reservoirs with regular jumps. *Neu. Compu.* 24, 1822–1852. doi:10.1162/NECO_a_00297

Sarankumar, R., Poongodi, P., and Umamaheswari, G. (2016). Adaptive phase adjustment and Channel Prediction strategies (APA-CPS) in MIMO-OFDM based cognitive radios. *Asian Journal of Information Technology* 15 (11), 1816–1824.

Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D., et al. (2019). Chimpanzee face recognition from videos in the wild using deep learning. *Sci. Adv.* 5, eaaw0736–9. doi:10.1126/sciadv.aaw0736

Son, W. S., and Han, D. S. (2021). "Analysis on the Channel Prediction accuracy of deep learning-based approach," in International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Jeju Island, Korea, April 2021, 140–143. doi:10.1109/ICAIIC51459.2021.9415201

Suganthan, P. N., and Katuwal, R. (2021). On the origins of randomization-based feedforward neural networks. *Appl. Soft Com.* 1, 107239–107245. doi:10.1016/j.asoc.2021.107239

Tan, R. H., and Wang, J. (2023). Channel equalisation method for optical fibre communication networks based on deep learning networks. *Laser J.* 44, 157–161. doi:10.14016/j.cnki.jgzz.2023.10.157

Trivedi, V. K., and Kumar, P. (2018). "BER performance of multi user scheduling for MIMO-OFDM and MIMO-SCFDMA broadcast network with imperfect CSI," in National Conference on Communications, Hyderabad, India, February 2018, 1–6. doi:10.1109/NCC.2018.8599983

Wang, G. (2022). A lightweight width learning system based on regularization and quantization techniques. *China Uni. Min. Tech*. doi:10.27623/d.cnki.gzkyu.2022.000296

Wang, J., Ding, Y., Bian, S., Peng, Y., Liu, M., and Gui, G. (2019a). UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems. *IEEE Access* 7, 96105–96112. doi:10.1109/ACCESS.2019.2929091

Wang, T., Wen, C.-K., Jin, S., and Li, G. Y. (2019b). Deep learning-based CSI feedback approach for time-varying massive MIMO channels. *IEEE Wirel. Commun. Lett.* 8, 416–419. doi:10.1109/LWC.2018.2874264

Xu, M., and Han, M. (2015). Adaptive elastic echo state network for multivariate time series prediction. *IEEE T Cy* 46, 2173–2183. doi:10.1109/TCYB.2015.2467167

Xue, C., Zhou, T., Zhang, H., Liu, L., and Tao, C. (2021). "Deep learning based Channel Prediction for massive MIMO systems in high-speed railway scenarios," in 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Virtual Conference, April, 2021, 1–5. doi:10.1109/VTC2021-Spring51267.2021.9448982

Yang, Y., Smith, D., Rajasegaran, J., and Seneviratne, S. (2020). Power control for body area networks: accurate Channel Prediction by lightweight deep learning. *IEEE Internet Things* 8, 3567–3575. doi:10.1109/JIOT.2020.3024820

Zhang, T., Wang, X., Xu, X., and Chen, C. L. P. (2019). GCB-net: graph convolutional broad network and its application in emotion recognition. *IEEE T Affect. Comput.* 13, 379–388. doi:10.1109/TAFFC.2019.2937768

Zhao, H., Zheng, J., Deng, W., and Song, Y. (2019). Semi-supervised broad learning system based on Manifold regularization and broad network. *IEEE T Circuits-I* 67, 983–994. doi:10.1109/TCSI.2019.2959886

Zhao, H. Q., and Lu, X. (2023). Width learning system based on generalised maximum correlation entropy criterion. *Signal Process.*, 1–7. doi:10.16798/j.issn.1003-0530.2023.11.005

Zhou, D. X. (2013). On grouping effect of elastic net. *Sta. Proba. Lett.* 83, 2108–2112. doi:10.1016/j.spl.2013.05.014

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 04 April 2024
DOI 10.3389/fenrg.2024.1378242

Check for updates

OPEN ACCESS

EDITED BY
Jinran Wu,
Australian Catholic University, Australia

REVIEWED BY
Zhesen Cui,
Changzhi University, China
Chanjuan Liu,
Shanghai Customs College, China

*CORRESPONDENCE
Akanksha Jain,
akanksha091091@gmail.com

RECEIVED 29 January 2024
ACCEPTED 05 March 2024
PUBLISHED 04 April 2024

CITATION
Jain A and Gupta SC (2024), Optimal placement of distributed generation in power distribution system and evaluating the losses and voltage using machine learning algorithms.
Front. Energy Res. 12:1378242.
doi: 10.3389/fenrg.2024.1378242

COPYRIGHT
© 2024 Jain and Gupta. 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 placement of distributed generation in power distribution system and evaluating the losses and voltage using machine learning algorithms

Akanksha Jain* and S. C. Gupta

Electrical Engineering, Maulana Azad National Institute of Technology, Bhopal, India

As the modern power system continues to grow in size, complexity, and uncertainty, traditional methods may occasionally prove insufficient in addressing the associated challenges. The improper location of distributed generation varies the voltage profile, increases losses and compromises network capacity. Machine learning algorithms predict accurate site positions, and network reconfiguration improves the capacity of the power system. The proposed algorithm is a hybrid of machine learning and deep learning algorithms. It cascades Support Vector Machine as the main model and uses Random Forest and Radial Neural Networks as classification algorithms for accurately predicting DG position. The non-linearity characteristics of the DG problem are directly mapped to the proposed algorithms. The proposed algorithm is employed on familiar test setups like the IEEE 33-bus and 69-bus distribution systems using MATLAB R2017 as simulation software. The R-squared (R2) values for all parameters yield a value of 1, while the MAPE values are minimal for the proposed cascaded algorithm in contrast to other algorithms of LSTM, CNN, RNN and DQL.

KEYWORDS

distributed generation (DG), optimal placement, distribution systems, machine learning, artificial neural network

1 Introduction

1.1 Background

As electricity demands grow, rather than solely relying on constructing new centralized power plants and transmission infrastructure, it is wise to integrate smaller-scale production units nearer to consumption hubs. Compact yet powerful generating units, commonly known as distributed generation (DG), have garnered increased attention due to their numerous advantages. Improper allocation of DG within the distribution system (DS) can have detrimental effects on the power system rather than providing benefits. To guarantee the best allocation of distributed generation (DG), several optimization investigations have been recorded in scholarly works.

The optimization of distributed generation (DG) is directed towards fortifying the reliability of the power network. The application of distributed generation has several advantages over conventional generation, such as technical, economic, and environmental, for the electric distribution company and the end consumers. Due to the development of

emerging technologies, an appropriate allocation involves various factors like reliability, power quality, voltage profile, power loss, and control stability of the electric grid.

Effective approaches for the recognition of power system control challenges have consistently been a focal point. To tackle this issue, machine learning and deep learning methodologies are utilized to predict occurrences, evaluate the multitude of variables and conditions, categorize them, and identify key features when managing power system control challenges across preventive, normal, emergency, and recovery situations. Machine learning methods play a crucial role in discerning patterns and structures within data, enabling the analysis, processing, prediction, and categorization of extensive pertinent to the evaluation of intricate power system issues. Numerous studies have explored machine learning algorithms, and the proposal introduces the cascaded machine learning algorithm (CML) specifically designed to optimize the siting and sizing aspects within distribution systems. Menke et al. (2019).

An approach centered on estimation for assessing the dimensions of DG (Distributed Generation) and its impact on the network, aiming to circumvent the cumbersome and obligatory application of load flow methods. Various machine learning techniques, including linear regression, artificial neural networks, support vector regression, K-nearest neighbours, and decision trees, have been leveraged for these estimations and applied across established test systems. Purlu and Turkay (2021).

1.2 Related work

The progression of distributed generation alongside the utilization of machine learning algorithms and various optimization methods contributes to enhancing the voltage profile with minimized losses within the transmission network. Recently, numerous authors have introduced a variety of algorithms, incorporating machine learning, deep learning, and several meta-heuristic functions, aimed at optimizing the allocation of distributed generation. In Essallah et al. (2019) The goal of the authors is to secure voltage stability amid load variations while simultaneously pinpointing ideal locations for distributed generation (DG) and determining their most suitable power capacities. Employing the PSAT MATLAB toolbox, they utilized this approach on the IEEE evaluation systems, encompassing the 33-bus and 69-bus configurations, revealing the method's efficacy, reliability, and robust performance. In the initial scenario, there was a noteworthy enhancement in the system's active power, exceeding its rated value by 50%. Additionally, both the active and reactive power loads saw a substantial increase of 50% in the subsequent scenario. In Sambaiah et al. (2019) the authors employed the Slaps Swarm Algorithm (SSA) for DG allocation. The concept is based on how slaps forage and navigate by swarming together in the water. In Liu et al. (2019) The total power losses are decreased by 90%, the cost is lowered by 21%, and the emissions are decreased by 67% for the 33-bus system.

To ascertain the optimal position and dimensions for DG units, the authors take into account multiple factors including voltage variation, line loss, and energy-saving benefits during the model selection process. This necessitates a comprehensive analysis to ensure accurate decision-making. This thorough evaluation ensures the optimal decision-making process. In Onlam et al. (2019) The authors utilized the Adaptive Shuffled Frog Leaping Algorithm (ASFLA) technique to carry out both network reconfiguration and DG installation. This approach was implemented across seven different scenarios for the electrical systems of IEEE 33-bus configuration and 69-bus configuration. In the most favorable scenario, the 33-bus system achieved loss savings of up to 75.57%, while the 69-bus system experienced an impressive 84.90% reduction in losses. Moreover, the Voltage Stability Index (VSI) increased by approximately 35.45% and 40.82%, respectively, surpassing their base values. In Mohammadpourfard et al. (2019) The contribution generates valuable training data through a scenario generator, along with specific details outlining the Artificial Neural Network (ANN) architecture. The results unequivocally illustrate that the proposed approach effectively addresses the limitations observed in existing ANN methodologies, particularly those that are not suitable for grids with significant Distributed Generation (DG) penetration. In Chege et al. (2019) the authors employ the voltage stability index technique to determine where distributed generation and capacitors should be placed. By employing a hybrid evolutionary programming method, the search for the most suitable sizes of distributed generation and capacitors to be installed at the indicated locations is facilitated. When the placement was carried out using this technique, the lowest voltage values of the network rose from 0.9036 to 0.9400 per unit, while the minimum Voltage Stability Index (VSI) values increased from 0.6690 to 0.7841. In Yin et al. (2019) to determine the feasibility and efficiency of the model, it is essential to carry out a thorough validation process. Therefore, a comparison is made between the results obtained and those achieved through the utilization of genetic algorithms (GA), support vector machines (SVM), and particle swarm algorithms (PSO). The kernel machine learning obtains the results of the capacity choice of the DG that satisfy the target function by training the model. In Shaheen et al. (2019) The authors suggest that the Stochastic Fractal Optimization (SFO) algorithm showcases validity, accuracy, feasibility, and robustness, surpassing alternative simulation methods in addressing the optimum load power flow (OPF) issue. In Arif et al. (2020) the authors suggest the implementation of the Analytical Hybrid Particle Swarm Optimization (AHPSO) algorithm to assist in evenly distributing, progressively dispersing, centrally distributing, and randomly distributing loads. Based on simulation outcomes, the AHPSO algorithm showcases significantly accelerated convergence rates compared to the traditional Particle Swarm Optimization technique. Notably, it demonstrates substantial enhancements in convergence performance across various distribution systems. The achieved enhancements are significant, demonstrating improvements of 32%, 42.59%, 50.91%, and 55.56% for the distribution systems of IEEE 10-bus, IEEE 33-bus, IEEE 69-bus, and KEPCO, respectively. Notably, both the normal PSO and AHPSO algorithms yield similar results for the location and scaling of DG in every scenario. In Hassan et al. (2020) the authors suggest a methodology directed at addressing the concern of siting and determining the dimensions of distributed generation units fueled by sustainable energy sources. Morales et al. (2020) A procedure is described to identify the current flow along

the distribution feeder (DF) while integrating distributed generation (DG). This method incorporates decision trees (DT), empirical decomposition (ED), and support vector machines (SVM) to achieve the intended objective, delineating the steps of the methodology. In Ismail et al. (2020) The authors introduce an algorithm designed for power systems utilizing deep learning principles.

This algorithm achieved an exceptional detection rate of 99.3% and demonstrated a minimal false alarm rate of only 0.22% by incorporating specific details related to PV generation, data, and interpretations extracted from SCADA systems. In Farh et al. (2020) an inventive technique, the Crow Search Algorithm-based Auto-Drive Particle Swarm Optimization (CSA-PSO) technique, is employed to optimize the setup, scale, and quantity of [unclear] systems with the goal of minimizing overall expenses and reducing power losses. In Tran et al. (2020) seven different scenarios were used to evaluate the proposed Stochastic Fractal search (SFS) algorithm on 33, 69, 84, 119, and 136 buses distribution networks. When the results from SFS were contrasted with those from other methodologies, it became clear that SFS produced more effective solutions. The SFS algorithm has the potential to search heuristics for solving issues quickly and with high-quality solutions. In Agajie et al. (2020) the application of the development of a Grid-based Multi-Objective Harmony Search Algorithm (GrMHSA) to improve the sizing and placement of Distributed Generation units ensures the enhancement of the overall system performance. In Admasie et al. (2020) the Intelligent Detection Method (IIDM) integrates a sophisticated approach termed Feature-based Wolf-Optimized 113. Artificial Neural Network (GWO-ANN) using an intrinsic mode function (IMF). Through MATLAB simulations, the IIDM's efficacy is highlighted in achieving high classification accuracy, computational efficiency, and robustness against noise interference in measured voltages. In contrast to directly resolving models under varying operational conditions, this method addresses the volt-VAR optimization (VVO) challenge in unbalanced distribution networks employing an advanced Deep Q-Network (DQN) framework. Its effectiveness is evaluated on IEEE 13-bus and 123-bus networks with imbalanced characteristics. In Haider et al. (2021) the application of the algorithm demonstrated significant efficacy in mitigating voltage fluctuations and reducing power losses within the distribution network. MATLAB software was employed to estimate the efficiency of this technique on the radial distribution network of the IEEE-33 bus system. Through the integration of a hybrid strategy named Enhanced Wolf Optimizer and Particle Swarm Optimization (EGWO-PSO), it became viable to attain optimized placement and sizing, resulting in an overall enhancement of the system's performance. This heuristic approach, EGWO, draws inspiration from wolf behaviors and presents a hybrid methodology that converges rapidly without being limited by local best practices Venkatesan et al. (2021). In Ogunsina et al. (2021) The authors employed the ant colony optimization (ACO) algorithm identifying the Optimal Positions and quantities for distributed resources within a power network. Bajaj and Singh (2021) The authors introduced a multi-objective method integrating multiple performance index constraints and compared its outcomes with hosting capacity augmentation techniques. The method involves the ENLPCI (Extended Nonlinear Load Position-based APF Current Injection) technique, facilitating the identification of optimal bus locations for Active Power Filter (APF) placement and the determination of required APF quantities.

The algorithm underwent performance evaluation through tests conducted on the IEEE test system with 69-bus. The outcomes highlighted the considerable impact of solar fluctuations on the best positioning and dimensions of Active Power Filters (APF)as part of the Optimal Placement and Sizing (OPAS) approach. Lakum and Mahajan (2021). In Lakum and Mahajan (2019) The authors introduced a novel technique called Non-Linear Load Position-Based Current Injection (NLPCI) aimed at strategically positioning the Active Power Filter (APF). The process involves the identification of suitable bus locations while considering the presence of Distributed Generation (DG) and excluding linear loads. To reduce the APF's size, the algorithm utilizes a voltage threshold of 5%, employing the Total Harmonic Distortion method. The implementation of this proposed algorithm notably enhanced system performance indices, increasing the Hosting Capacity (HC) across various test systems, including the 59-node Egyptian test system, 135-node Brazilian test distribution network, additionally, the analysis includes the IEEE 33-node and IEEE 69-node test power networks. Moreover, using the NR (Newton-Raphson) and AOA (Artificially Optimized Algorithm) enhanced the HC of the 59-node and 135-node test systems by 38.832% and 72.895%, respectively. Ali et al. (2021) The use of the Meta-heuristic Rider Algorithm (ROA) aids in evaluating the optimal dimensions and positions for Distributed Generation (DG) units, which are driven by sustainable energy resources like biomass-based generation, wind turbines (WT), and photo voltaic (PV) systems within distribution networks. Khasanov et al. (2021) The authors ingeniously employed the coyote technique to efficiently determine the positioning and capacity of distributed generation (DG) within radial networks. This approach demonstrated superior performance compared to other methods like SFO, SSA, and COA in making optimal DG decisions. Pham et al. (2021). The authors assessed their proposed Modified Sequential Switch Opening and Exchange (MSSOE) algorithm using industry-standard distribution systems like IEEE 33-node, IEEE 69-node, and IEEE 119-node. The findings show that MSSOE surpasses existing algorithms, delivering a superior global solution with reduced computational time. Vannak Vai et al. (2021). The study incorporated the Firefly Algorithm (FA), a meta-heuristic technique, within the IEEE 33-bus distribution system as part of its methodologies. This method includes the application of a mixed probabilistic model to manage both output power uncertainty and the system load. Naguib et al. (2021). The research integrates hybrid particle swarm optimization (PSO) algorithms, these methods incorporate chaotic maps and adaptive acceleration coefficients to identify the optimal configuration and dimensions of photovoltaic (PV) systems connected to an electric grid. Belbachir et al. (2021). In Bajaj et al. (2020) A methodology inspired by the Analytical Hierarchy Process (AHP) is introduced for evaluating Power Quality (PQ) in modified distribution power systems undergoing changes. In Kushal et al. (2020) The authors have introduced a decision-making methodology that uses the Analytical Hierarchy Process (AHP) to assess various distributions of Photovoltaic (PV) and Battery Energy Storage Systems (BESS) resources to load buses., considering various scenarios focused on cost and resilience enhancement. The results indicate that this approach offers

valuable guidance for planners in making informed investments. Additionally, they introduced a hybrid technique merging clustering with particle swarm optimization to optimize Distributed Generation (DG) allocation. This algorithm clusters buses within the distribution network and selects the most suitable cluster for DG allocation, thereby reducing the number of viable production buses. In the optimal scenarios, the algorithm reduces test feeder losses by 69%, 86%, and 90%, while considering uncertainties in three different scenarios, achieving reductions of 39%, 53%, and 55% respectively Eyüboğlu and Gül (2021). The revised method for updating control variables of the initial slime algorithm led to the development of a new solution. As a result, the proposed algorithm demonstrates a notable improvement in reducing voltage deviation to 1.4779 Pu and power loss to 78.88%. In contrast, alternative algorithms provided results within the range of 69.10%–78.87% for loss and 1.5759 to 1.4996 Pu for voltage deviation Pham et al. (2022). The authors incorporated wind generation into the distribution system, resulting in enhanced bus voltages and reduced active power losses. Specifically, nodes 19 and 12 experienced voltage increases of 2.2% and 3.26%, respectively. Consequently, in the IEEE-19 bus system, losses reduced from 13.54 kW to 6.36 kW, while in the IEEE-25 bus system, losses decreased from 150 kW to 95.30 kW Routray et al. (2021). The Loss Sensitivity Factor plays a pivotal role in determining the search space for Distributed Generation (DG) sites. Employing an analytical technique, the initial DG sizes are calculated based on a specific interpretation. Selim et al. (2021). In Rathore and Patidar (2021) The authors introduce a deep learning-based algorithm designed for effectively placing and arranging DG units in optimal locations. This algorithm aims to enhance voltage characteristics and minimize losses within the system. In Dheeban and Selvan (2021) The authors utilized the Thevenin Equivalent Impedance method to simulate the distributed generation (DG) source on the high-voltage side of the transformer. The previous sections demonstrate various methods for determining distribution system voltages, including distribution load flow for unbalanced systems, symmetrical components, mesh, and nodal analysis. An enhanced algorithm, incorporated into an adaptive interference controller, was then applied to the proposed PV-UPQC (Photovoltaic - Unified Power Quality Conditioner) system. This advanced fuzzy-model-based controller enhances the system by analyzing system parameters and facilitating the generation of reference current Zhang et al. (2020). Mo and Sansavini (2019) The exploration involves utilizing Optimal Power Flow (OPF) to calculate the suitability of Energy Not Served (ENS) and Operation and Maintenance (O&M) costs for Distributed Generation (DG). This process tends to overestimate unreliability and costs. To address this issue, the simulation results provide guidance for decision-making in managing, maintaining, and planning Distributed Generations. The objective is to mitigate the effects, minimize O&M expenses, and decrease energy loss, ensuring more efficient and reliable systems Zhao et al. (2019). The paper undertakes a comparative analysis between two operational strategies and multiple islanding strategies for coordinating dispersed power and storage devices. The study's results demonstrate the considerable enhancement in the dependability of the distribution system caused by implementation of the proposed dynamic strategy. This enhancement maintains the steady functioning of vital loads within the system. Furthermore, it deeply explores fault detection and protection techniques employed in distributed networks alongside distributed generation. This paper uniquely contributes by offering a comprehensive understanding of diverse fault detection strategies, emphasizing the operational and communication methodologies of Distributed Generation (DG), different parameters involved, and their respective limitations Nsaif et al. (2021). In Lotfi (2022) The method's effectiveness is evaluated on both a 95-node and a 136-node test system. The results indicate significant improvements compared to the base standards. In the first system, energy losses reduced by 11%, operational costs by 25.5%, and Energy Not Served (ENS) by 5% through the optimization of the distributed units and shunt capacitors values. Moreover, in the second system, when employing the Time-of-Use mechanism to determine DG units, there was a remarkable decrease of 29% in energy losses, 65% in operational costs, and 7% in ENS compared to the base values. In Bhusal et al. (2022) The method was tested on diverse systems, including the IEEE 123-node distribution setup and an actual 240-node system in the United States. Findings demonstrate that the approach effectively identifies coordinated attacks with remarkable accuracy of up to 99.9%. The research employs different neural networks, encompassing Convolutional Neural Networks (CNN), Residual Neural Networks, and Multi-Layer Perceptrons (MLP) to evaluate the effectiveness of this method. Additionally, the method contributes significantly to precisely detecting the attack type within a DG, distinguishing between additive or deductive attacks. This distinction provides operators with explicit data, enabling them to undertake corrective measures effectively. The Gauss-Seidel (GS) method is applied for load flow analysis, offering a simple iterative approach to solving load flow equations when partial derivatives are not required Alnabi et al. (2022). A completely distributed energy trading system based on machine learning was proposed. Design skeletons for a DC grid energy management system are designed by developing an experimental strategy and assessing the efficacy in comparison to other available options. In Yılmaz et al. (2022) the pyramidal algorithm with wavelet transform is a novel feature extraction strategy used in the proposed hybrid machine learning method (UWT). With an accuracy rate of up to 99.59%, the UWT SGBT approach effectively groups data according to mathematical and real data results. Furthermore, this approach has been extensively tested in noisy environments, and the pyramidal UWT-SGBT method demonstrates superior noise immunity compared to other machine learning methods that utilize wavelet transform (WT)-based techniques. In Gawusu et al. (2022) The study delves into the exploration of patterns and trends in dispersed generation through data mining (DG). While still at an early stage, this research provides valuable insights into ongoing research trends and patterns. These findings would be particularly beneficial and of significant interest to researchers actively engaged in the Distributed Generation field. In Bhadoriya et al. (2022), As the load increases, distribution network losses can cause voltage deviations and compromise stability. In response, the Transmission System Operator (TSO) has developed an optimization solution for the allocation of Distributed Generation (DG) that meets various equality and inequality requirements. The goal is to mitigate reduce voltage fluctuation, decrease active power dissipation and guarantee voltage stability. Significantly, both the IEEE 33-bus and 69-bus distribution systems have experienced considerable decrease

in active power dissipation, reaching percentages of 94.29% and 94.71%, respectively. In Pereira et al. (2016), a combined strategy that merges tabu search with genetic algorithms (GA) has been proposed for the incorporation of distributed generation (DG) and capacitor banks into distribution systems (DS) to improve overall system efficiency. In Battapothula et al. (2019), an innovative optimization technique, utilizing a hybrid shuffled frog leap teaching and learning algorithm, has been suggested and implemented to determine the most efficient placement and dimensions of electric vehicle rapid charging stations and distributed generation (DG) units within distribution systems (DS). In Bo et al. (2020), a hybrid load forecasting model has been developed by integrating data preprocessing techniques, forecast algorithms, and weight identification theories. An approach utilizing wavelet decomposition and quadratic gray neural network, in conjunction with the enhanced Dickey-Fuller test, Li et al. (2017), dynamic model selection relying on Q-learning Feng et al. (2020), a method employing boosting-based multiple kernel learning Wu et al. (2020), hybrid method that merges convolutional neural networks (CNN) with long short-term memory (LSTM) for deep learning (DL). Alhussein et al. (2020), a DL method Hong et al. (2020), and a hybrid model clustering with feed forward neural networks (FFNN) Panapakidis et al. (2020) have been proposed for STLF. A proficient forecasting model is introduced, incorporating a feature extraction module that merges variational mode decomposition (VMD) with a variational auto encoder (VAE) Yang et al. (2022). The project entails sharing data, which necessitates disclosing the private information of the participants. To tackle this concern, the authors integrated variational mode decomposition (VMD), the federated k-means clustering algorithm (FK), and SecureBoost into a unified algorithm, named VMD-FK-SecureBoost. Yang et al. (2023).

1.3 Research Gap

A machine learning algorithm was recently employed in several data and flow-oriented applications for prediction and selection. The non-linear machine learning algorithm mapped the objectives and multiple constraints of DG placement and sizing. Machine learning provides several algorithms in the domain specified, such as supervised, unsupervised, and reinforced learning. The reported survey suggests that most authors employed supervised learning-based algorithms such as support vector machine (SVM), K-nearest (KNN), decision tree (DT), extreme learning (ELM), and many more hybrid algorithms. The limitations of feature selection and mapping of machine learning algorithms are overcome with deep learning algorithms.

As evident from the literature review, it is commonly acknowledged that the load is typically presumed to remain constant, and the DG output can be regulated during DG allocation. However, in real-world scenarios, both loads and DG output fluctuate continuously. This variability poses challenges when calculating losses and other parameters using power flow-based algorithms, making the process cumbersome and time-intensive.

Several methodologies focus solely on active power injection, specifically for distributed generation (DG) with a unity power factor. Some techniques are designed for allocating a single DG unit, neglecting environmental considerations. Certain network constraints may go unnoticed, and economic factors are overlooked in certain articles. Furthermore, certain methods exhibit excessively long computational times, and the outcomes derived from certain approaches may not be optimal.

1.4 Paper contributions

The proposed algorithm effectively addresses issues such as voltage imbalance, line loading, and power losses within the distribution system. The frequently employed algorithm for Distributed Generation (DG) allocation is the support vector machine (SVM). Recently, several authors employed convolutional neural networks (CNN) for DG allocation. The proposed algorithm offers a significant advantage by overcoming limitations present in Artificial Neural Networks such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), including Long Short-Term Memory (LSTM) and Deep Q-learning (DQL) algorithms, while accumulating various benefits.

The specific contributions of the work proposed are as follows:

1. Estimation of generation unit sizes, power losses within the distribution network, and minimum voltage (without power flow estimation)utilized in the test scenarios of IEEE 33-bus and IEEE 69-bus distribution systems.
2. Evaluation and comparison of RNN, CNN, LSTM, and DQL algorithm performances on similar test systems, utilizing metrics like Mean Absolute Percentage Error (MAPE) and R2.

1.5 Paper organization

The forthcoming sections of this manuscript will follow this structured approach:

Section II: Methodology description detailing DG allocation and sizing.

Section III: Experimental analysis, presenting the results.

Section IV: Comprehensive overview of the entire study.

Section V: Exploration of potential future avenues in this field of research.

2 Methodology

The application of Machine Learning for predicting the optimal locations for distributed generation is explored in this study. The paper delves into four key algorithms: Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) under the Neural Network category, Deep Q-Learning (DQL) as a Reinforcement Learning method, and Long Short-Term Memory (LSTM) representing Supervised Machine Learning.

In a quest to enhance distribution system performance, a novel approach named the cascaded machine learning algorithm is introduced. This algorithm combines Random Forest, Radial Neural Network, and Support Vector Machine (SVM). Its

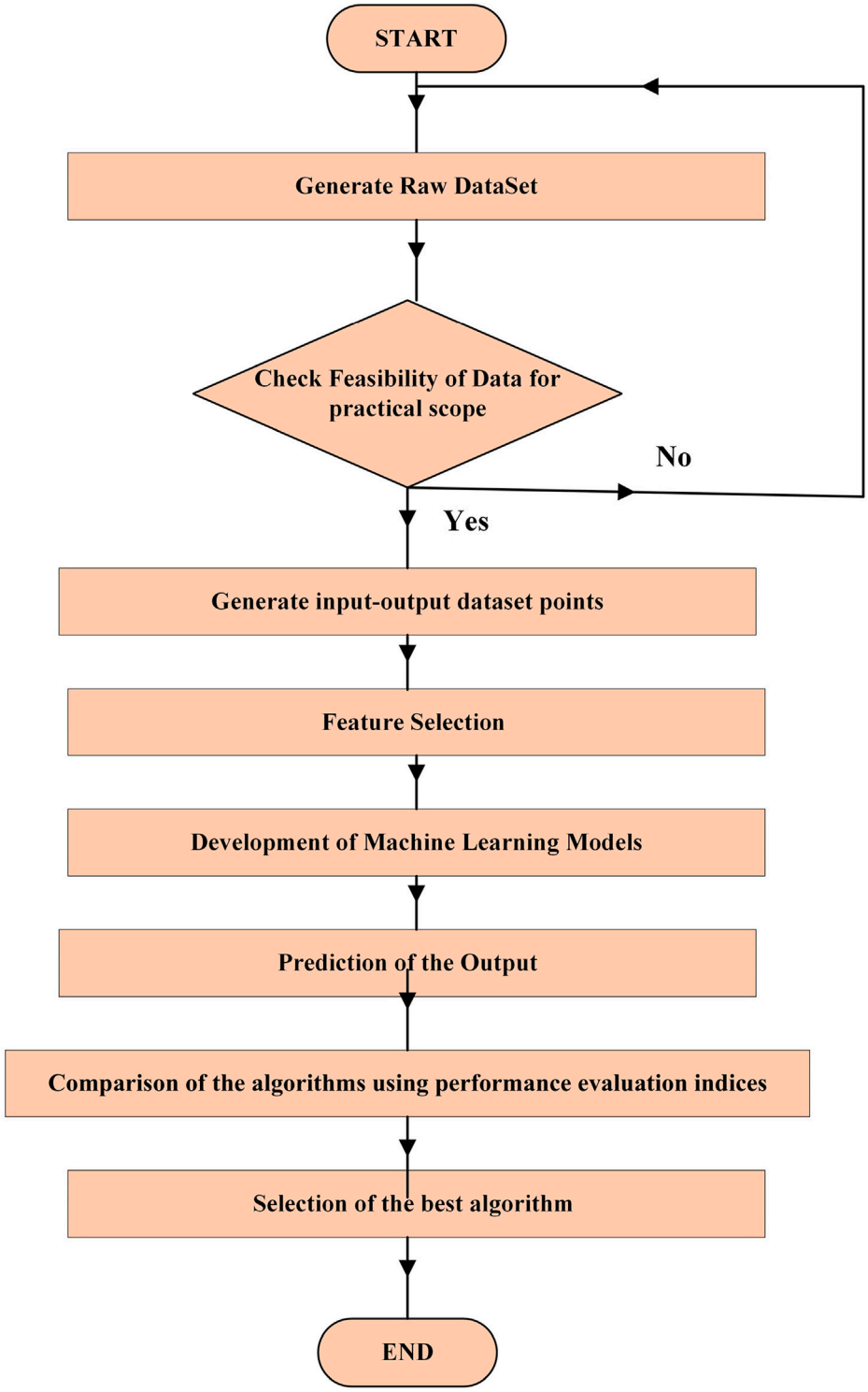


FIGURE 1
Flowchart for the process.

primary goal is to enhance the placement of Distributed Generation (DG)with the intention of minimizing power dissipation within the distribution system. The investigation comprehensively evaluates the impact on various aspects such as Distributed Generation power injection, minimum voltage for, and both active and reactive electrical losses. The study employs test systems based on the IEEE 33-bus and 69-bus configurations. In the method presented, the division of data is random, with a ratio of 80:20 for training and validation.

The process encompasses various key steps, including feature selection (input), training and testing phases, as well as validation. Within this framework, multiple ML models—specifically RNN, CNN, Proposed model, DQL, and LSTM are constructed to facilitate

TABLE 1 DG sizing estimation analysis.

Distribution for test system	Algorithm applied	R2	MAPE [%]
33-bus	RNN	0.9785	0.0247
	CNN	0.9678	0.0666
	LSTM	0.9885	0.0657
	DQL	0.9857	0.0013
	Proposed	1	0.0012
69-bus	RNN	0.9851	0.0300
	CNN	0.9845	0.0578
	LSTM	0.9852	0.0789
	DQL	0.9734	0.0025
	Proposed	1	0.0017

TABLE 2 Active power loss estimation analysis.

Distribution system for test	Algorithm applied	R2	MAPE [%]
33-bus	RNN	0.7078	26.0063
	CNN	0.9685	1.4671
	LSTM	0.9865	26.0765
	DQL	0.7846	1.4854
	Proposed	1	0.1917
69-bus	RNN	0.5978	26.5257
	CNN	0.9871	1.4745
	LSTM	0.9784	26.1838
	DQL	0.9257	1.1570
	Proposed	1	0.1918

predictions. The validation of the final model is conducted using the method of hours to ensure accurate predictions for both the testing setups of IEEE 33-bus and 69-bus systems. Furthermore, the entire process is elucidated through a flowchart in Figure 1, outlining the comprehensive process.

2.1 Machine learning algorithms applied

The Machine Learning models were constructed using generated data sets to facilitate training. Regression-based training algorithms, including Support Vector Machine (SVM), CNN, and CML, were utilized. Subsequently, their respective R2 scores and MAPE were computed to predict reactive power losses and minimum busbar voltages. The outcomes of Mean Absolute Percentage Error (MAPE) and R-squared (R2) values for all five models are showcased in Tables 1–4. These models allow the extraction of output parameters for forthcoming input data. The study applied the following machine learning algorithms:

2.1.1 Support vector Machines

Support vector machines (SVM) were created for empirical data as regression models and binary classifiers as prediction models. The operation of a support vector machine maps input data onto a two-dimensional space via a hyperplane. The significance of this hyperplane lies in its ability to classify data into distinct categories. The effectiveness of support vector machines relies heavily on the appropriate choice and maximization of hyperplanes. Optimization enhances the optimal scope of the hyperplane, leading to improved predictions by the support vector machine. The equation for hyperplane is given as Eq. 1

$$\min \frac{1}{2}|w|^2 + C\left(\sum_{i=0}^{1} T_i\right) \tag{1}$$

Here data to y_i $(w.x_i + b) \geq 1\text{-}T_i$, $T_i \geq 0$ for all i

where y_i stands in for the class label, which is either +1 or −1, and x_i represents the i^{th} example. The issue's dual form is used to solve it, given in Eq. 2

TABLE 3 Reactive power loss estimation analysis.

Distribution system for test	Algorithm applied	R2	MAPE [%]
33-Bus	RNN	0.9952	2.3209
	CNN	0.9973	1.4870
	LSTM	0.6009	25.0065
	DQL	0.9570	8.1056
	Proposed	1	0.1718
69-bus	RNN	0.9952	2.330
	CNN	0.9973	1.7835
	LSTM	0.588	25.5358
	DQL	0.9558	8.2616
	Proposed	1	0.1906

TABLE 4 Minimum busbar voltage estimation analysis.

Distribution system for test	Algorithm applied	R2	MAPE [%]
33-bus	RNN	0.9933	0.0412
	CNN	0.9957	0.0477
	LSTM	0.9983	0.0246
	DQL	0.9997	0.0079
	Proposed	1	0.0011
69-bus	RNN	0.9933	0.0430
	CNN	0.9957	0.0497
	LSTM	0.9977	0.0400
	DQL	0.9996	0.0096
	Proposed	1	0.0018

$$\max L_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_{iy_j} (x_i{}^T x_j) \quad (2)$$

Where $0 \le \alpha_i \le C$ for all i, $\sum_i \alpha_i y_i$

2.1.2 Convolutional neural network (CNN)

A Convolutional Neural Network (CNN) is structured with multiple layers that serve distinct roles in data processing. This architecture comprises input layers, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. The effectiveness of CNN classifiers in data classification and detection is attributed to the diverse capabilities of these layers.

The convolutional layer, through various window sizes, extracts different feature information from the data matrix by employing convolution kernels. Utilizing this convolution operation enables parameter sharing, wherein the same weight and offset are shared across the network. As a result, this parameter-sharing technique substantially diminishes the total number of parameters required by the neural network, thereby amplifying its efficiency.

After the convolutional layer, the pooling layer implements diverse sampling techniques to down sample the feature map, typically using mean or maximum values within specified window regions. This down-sampling process reduces the size of the features, effectively diminishing the impact of redundant information and aids in managing computational complexity.

Consider that the input features of CNN are a map of layer x, M_x (M_0 = F)

The convolutional process can be expressed in Eq. 3 as Li et al. (2017); Wu et al. (2020)

$$M_x = f(M_{x-1} \otimes W_x + b_i) \quad (3)$$

Here W_x is the convolutional kernel weight vector of the x layer,

Symbol ⊗ represents convolutional approach,

b_i is the offset vector of x layer. F(x) represents the activation function. By employing diverse window values, the convolutional layer extracts varied feature information from the data matrix, utilizing different convolution kernels. Embracing the concept of 'parameter sharing,' the convolution operation involves using the same weight and offset for all convolution kernels, thereby significantly reducing the total number of parameters in the neural network. Subsequent to the convolutional layer, the pooling layer typically conducts feature map

sampling through various algorithms. The pooling layer can be expressed as follows, where

M_x is the input, and M_{x+1} is the output of the pooling layeras given in Eq. 4

$$M_{x+1}\ \text{subsampling}\,(M_x) \tag{4}$$

The sampling criterion typically involves selecting the mean or maximum value within the window region. The main objective of the pooling layer is to reduce the size of the features, thereby mitigating the impact of redundant features on the model.

2.1.3 Recurrent neural network

The Recurrent Neural Network (RNN) distinguishes itself as a specialized variant of an artificial neural network designed particularly for analyzing sequential time series data. One of its key advantages lies in its ability to process data in a sequential manner through the establishment of signal pathways both forward and backwards. This is accomplished by creating internal loops within the network, enabling connections among hidden components.

RNNs, due to these internal connections, are adept at leveraging information from past data to predict future data points, a capability that becomes particularly useful when exploring temporal correlations among diverse data sets. These networks excel in capturing and understanding sequential patterns within data sequences, making them highly suitable for tasks involving time-dependent data analysis and predictions. Wang et al. (2019) Wang et al. (2020).

2.1.4 Long short-term memory

The inception of Long Short-Term Memory (LSTM) was driven by a specific aim: to address the challenge of vanishing gradients that conventional Recurrent Neural Networks (RNNs) encounter when handling long-term dependencies. Unlike the simple chain-like structure of a regular RNN consisting of a series of recurrent units forming a single layer, the architecture of LSTM's hidden layers is more intricate and involves a sequence of repeating modules.

LSTM (Long Short-Term Memory) incorporates gate mechanisms and memory cells within each hidden layer, forming a memory block with essential components: an input gate, an output gate, memory cells, and a gate. Each element fulfils specific functions in regulating information flow within the network. The input gate controls which data to preserve in the memory cell, while the output gate determines when to transmit this information to the next layers. The forget gate, on the other hand, aids the network in discarding unnecessary or outdated information, essentially resetting the memory cells.

Crucially, LSTM employs multiplicative gates that facilitate the retention and access of information over longer time intervals. By strategically incorporating these gates, LSTM significantly mitigates the problem of vanishing gradients, ensuring the network can effectively capture and retain relevant information over extended sequences, thereby enhancing its ability to handle long-term dependencies in data. Hu et al. (2020) Qiao et al. (2020).

2.1.5 Deep Q learning

Deep Learning holds significant promise in transforming power load forecasting methodologies. Its impact on the energy sector's data processing methods is noteworthy. Artificial Neural Networks (ANNs) specifically engineered for this objective include multiple layers positioned among the input and output strata. These layers enable the seamless transmission of information forward across the network through a process known as feed-forward propagation. This architecture enables the ANN to process forecasting data effectively, contributing to the growing acceptance and application of deep neural networks within the energy sector. Eyüboglu and Gül (2021).

2.1.6 Proposed Algorithm

The Cascaded Machine Learning algorithm enhances data training and minimizes network errors by integrating three supervised learning algorithms: Random Forest, Radial Neural Network and Support Vector Machine (SVM). Within this context, the Support Vector Machine serves as the central processing model. Meanwhile, the Random Forest and Radial Neural Network serve as classification algorithms. The Random Forest classifier operates for variable feature selection from DG data, whereas the Radial Neural Network is responsible for predicting the optimality of the DGs. This amalgamation enables improved data processing and accurate prediction within the network.

The processing of the cascaded machine-learning algorithm is described in Eq. 5 as

$$M_c\,(x) = \text{sgn}\,m_1\,(x))\text{sgn}\,(m_2\,(x)\,\text{sgn}\,(mn\,(xn))) \tag{5}$$

Here $M_c\,(x)$ is the multiple-stage function of the support vector machine. The processing of multiple stages is carried out on the feature data of DGs on distributed networks. For the optimality of DGs, this employs support vector machine function SVM (f_{svm}) to separate the hyperplane of optimal and non-optimal data, as given in Eq. 6.

$$F\,(x) = \begin{cases} -1 & \text{if } M_c\ (x) < 0 \\ \text{sgn}\ (f_{svm}\ (x)) & \text{if } M_c\ (x) \geq 0 \end{cases} \tag{6}$$

The selection of features data from DGs matrix is given in Eq. 7 as

$$Rf\,(x) = (P1, \ldots, Pm] \tag{7}$$

Here $p_1, \ldots, p_n$ are the vectors of selected DGs matrix data.

Estimate the entropy of class by variable selector,as mentioned below in Eq. 8

$$LX\,(x) = (E_{i1}, \ldots, E_{ik}\,] \tag{8}$$

Here E is entropy of variables of DGs matrix.

The training set is derived pairwise (s, $[E_1, \ldots, E_k]$)

Where s represent the input variable and $[P_1, \ldots, P_k]$ the target variable.

$$e_j = \begin{cases} 1, \text{if } j = k \\ 0, \text{otherwise} \end{cases} \tag{9}$$

Eq. 9 signifies multiple support vectors. The optimal process of the variable is selected, and then the process pattern on the radial neural network call adjusts the weight factor of the cascading process.

$Mc \in ej^{d\sim} \leftarrow$ mapping of optimal data in kernel function.

Call kernel function as described in Eq. 10

$$k\left(x_i, x_j\right) = \exp\left(\frac{\|\ x_i - x_j\ \|^2}{\gamma}\right), \gamma \in R \tag{10}$$

End if.

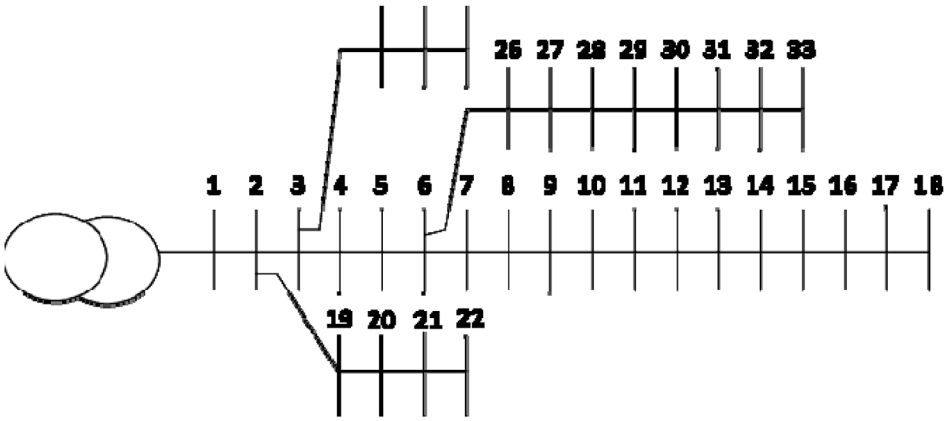


FIGURE 2
IEEE 33-bus test system.

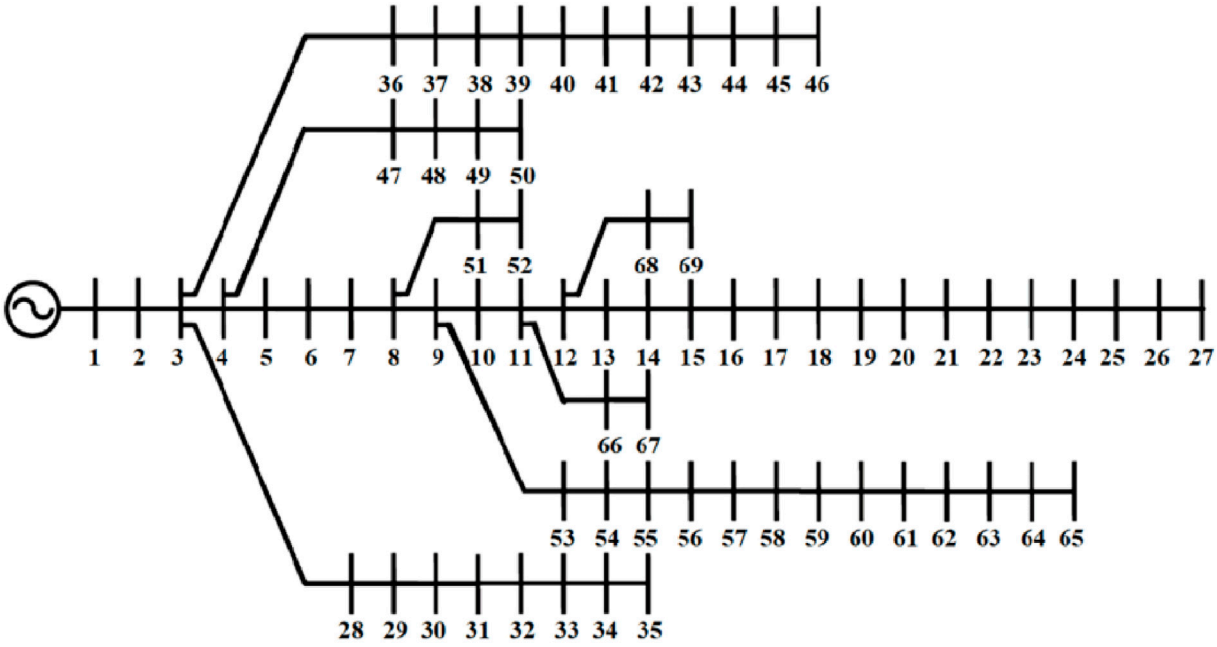


FIGURE 3
IEEE 69-bus test system.

3 Simulation and results analysis

The simulation of the entire process is conducted through MATLAB R2017 software. It includes information regarding DG power injection, bus bar voltages, load variations, as well as both active and reactive power losses. The evaluation of these parameters is conducted using the IEEE 33-bus and IEEE 69-bus distribution networks as case studies. Figure 2 depicts the single-line diagram for the 33-bus distribution system, and Figure 3 illustrates the diagram for the 69-bus distribution system. This simulation methodology enables a comprehensive examination and assessment of the system's performance across various scenarios and conditions. Arif et al. (2020). The power load data obtained varies based on the time of day and seasonal changes. These data can be found in the GitHub repository Lotfi (2022). The assessment of the algorithms' performance involves the utilization of MAPE and R-squared indices. The results in the upcoming figures illustrate the distributed generation (DG) power injection, active power losses, reactive power losses, and minimum busbar voltage for both the IEEE 33-bus and 69-bus networks, respectively. The gathered R2 and MAPE values used for prediction are presented in Tables 1–4. Each case is thoroughly explored in the subsequent section, systematically analyzing the results and observations.

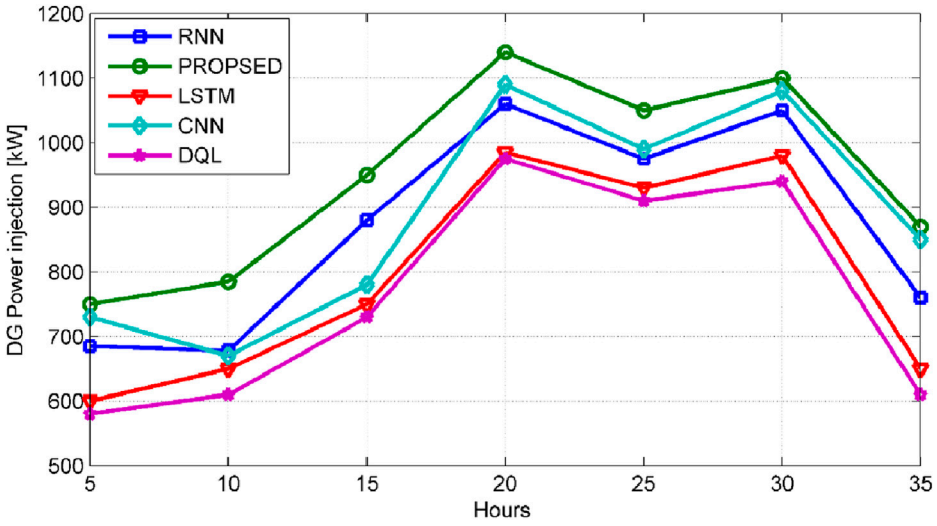


FIGURE 4
A comparative analysis of distributed generation (DG) sizing predictions for a 33-bus distribution system.

3.1 DG power injection

Figures 2, 3 illustrate the single-line diagram of the 33-bus and 69-bus distribution systems, respectively. Distributed generation (DG) sizing is determined through the utilization of normalized load variation (NLV).

P_{Dtt} = 2566.1833 x Load Level - 53.5766 for 33-bus distribution system.

P_L = 286.3641 x Load Level - 97.0291 for 33-bus distribution system.

Figure 4 depicts the optimal power injection at different intervals of scheduled working hours. The methodology for power injection prediction encompasses the utilization of diverse deep learning algorithms, including long short-term memory networks (LSTM), convolutional neural networks (CNN), deep Q-learning (DQL), recurrent neural networks (RNN), and the defined algorithm. Comparatively, the sequential algorithm of RNN demonstrates superior performance when compared to the CNN algorithm, leading to a notable 7% enhancement in prediction accuracy. Additionally, the proposed algorithm exhibits superiority over existing algorithms in terms of prediction ratio, which falls within the range of 5%–7%.

P_{Dtt} = 1827.2793 x Load Level - 4.711 for bus 69 distribution system

P_L = 318.0325 x Load Level - 108.6808 for bus 69 distribution system.

Figure 5 depicts the optimal power injection at different intervals of scheduled working hours. The 69-bus system exhibits a diverse prediction compared to the 33-bus system. The figure illustrates the deep learning-based approaches employed for the assessment of optimal DG prediction to enhance the distribution systems. The proposed algorithm for DG optimal prediction reached the maximum prediction of the actual value of DG sizing. The proposed algorithm overcomes the limitations of LSTM, RNN, CNN, and DQL. The RNN and LSTM algorithms suffer from training error rates; however, the prediction rate of RNN and LSTM is less than the actual value of prediction. The CNN algorithm is better than RNN and LSTM. The overall improvement in the prediction ratio of DG sizing is 4%–8% compared to existing algorithms.

3.2 Active power loss estimation

Figure 6 displays the predictions for active power losses in the 33-bus systems based on normalized load variation (NLV). We employed deep learning based algorithms such as RNN, LSTM, CNN, DQL, and the proposed algorithm to predict active power loss. The variation of algorithms employed in deep learning is estimated in different hours of optimal DG sizing. The LSTM algorithm exhibits inferior performance compared to both CNN and RNN algorithms. The LSTM algorithm suffers from the problem of training errors in prediction data. The conventional RNN algorithms are better than LSTM. The cascaded proposed algorithm overcomes the training errors and improves the prediction of active power loss.

Figure 7 illustrates the forecasts for active power loss by utilizing the normalized load variation (NLV) of 69-bus systems. The employed DQL algorithm for deep learning is inferior to other algorithms for prediction. The primary concern lies in the sluggish training rate of Deep Q-Learning (DQL), which adversely affects the prediction rate of active power loss compared to CNN, RNN, and LSTM. To address this limitation, the proposed algorithm is introduced, markedly enhancing the accuracy of active power loss prediction and maximizing accuracy in forecasting actual values. In the context of 69-bus systems, optimal sizing results in a notable 5%–10% improvement in the predictive performance for active power loss.

3.3 Reactive power loss estimation

Figure 8 displays the forecasted reactive loss for the 33-bus systems, where the CNN algorithm utilized for deep learning shows

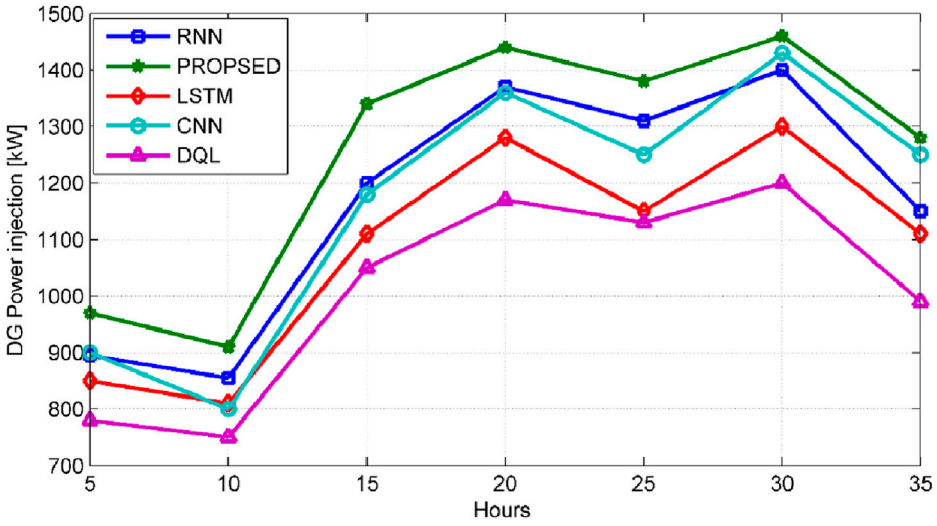


FIGURE 5
Comparing DG sizing predictions for the 69-bus distribution system.

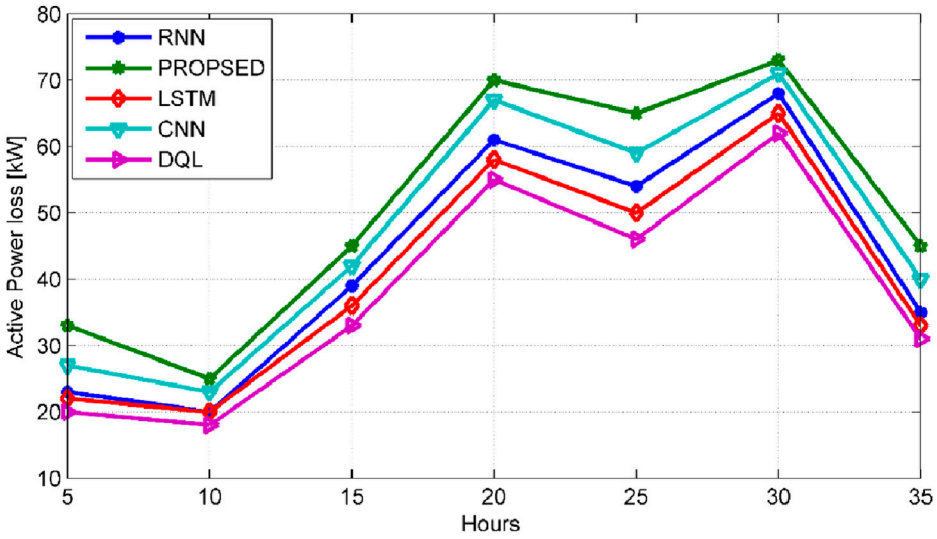


FIGURE 6
Comparative analysis of active power loss forecasts for a 33-bus distribution system.

inferior performance compared to other deep learning algorithms in prediction accuracy. The challenge primarily arises from the selection of the filter matrix for DG sizing data, resulting in errors and a reduction in active power loss prediction rates when compared to DQL, RNN, and LSTM. However, with the integration of the proposed algorithm, there's a notable enhancement in active power loss prediction and a substantial improvement in actual value prediction. For the 33-bus systems, achieving optimal sizing results in a notable enhancement of 7%–12% in the prediction ratio of reactive power loss.

Figure 9 showcases the predictions of reactive power loss based on actual values specifically for the 69-bus system. The predictive analysis utilized various deep learning algorithms, namely, RNN, LSTM, CNN, and DQL. The variance in prediction results stems from the training sequence applied to the DG sizing data, significantly influencing the outcome. The training process significantly shapes and exacerbates errors, consequently affecting the performance of RNN. Moreover, alternative deep learning algorithms like DQL, CNN, and LSTM show divergent predictions for reactive power loss. The implementation of the designated algorithm results in an enhancement of the predictive accuracy for reactive power loss by 2%–8% in comparison to the observed reactive power loss.

3.4 Minimum busbar voltage estimation

The precise estimation of Minimum Busbar voltages in the 33-node and 69-node distribution systems has been successfully

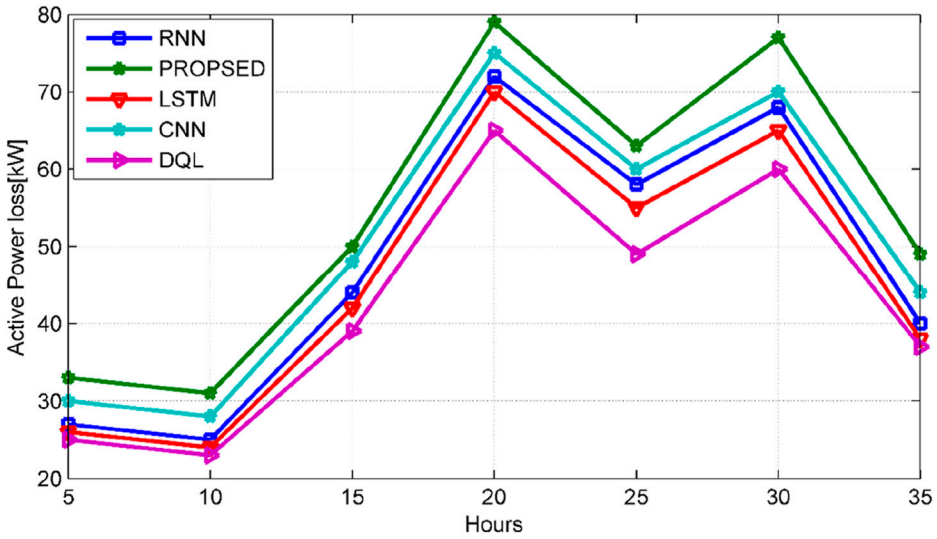


FIGURE 7
Comparative analysis of active power loss forecasts for a 69-bus distribution system.

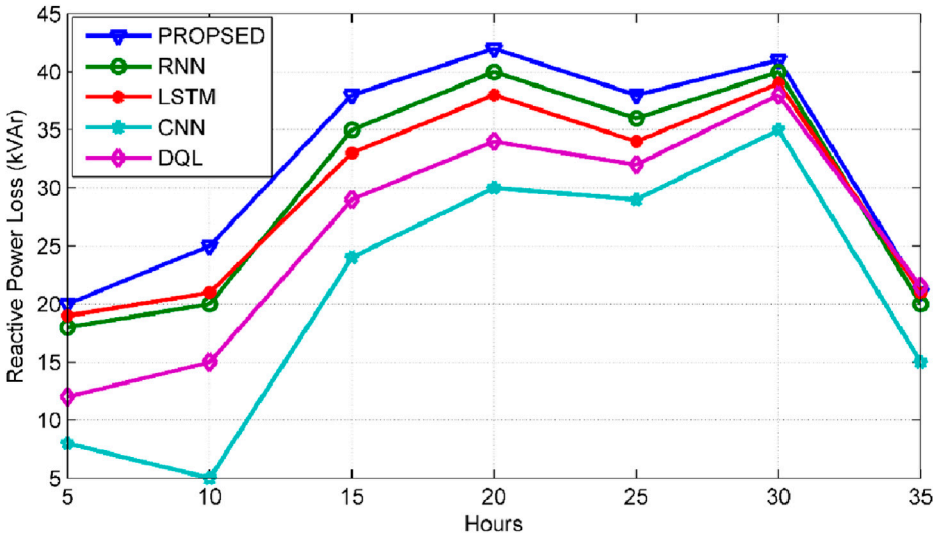


FIGURE 8
The predictive analysis of reactive power loss in the 33-bus distribution system.

accomplished by deploying the advanced Cascaded Machine Learning model. Figure 10 displays the prediction of minimum busbar voltages specifically related to the 33-bus distribution system. Various deep-learning algorithms of RNN, LSTM, CNN, and DQL were employed for minimum voltage prediction. Since busbar voltage is a time-series data, RNN and LSTM are more suitable for its prediction. However, due to slow training error rates, the LSTM's prediction rate is relatively lower compared to other deep learning algorithms. The proposed algorithm successfully addresses the issue of slow training error rates and notably enhances the prediction rate, improving the actual prediction rate by 5%.

Figure 11 depicts the forecast of minimum busbar voltage focusing on the 69-bus grid system. In this investigation, analogous to the 33-bus system, RNN, LSTM, CNN, and DQL models were employed to perform predictions for minimum voltage. With busbar voltage being a time-series data, RNN and LSTM are considered more effective for prediction. Nevertheless, due to slow training error rates, the DQL's prediction rate is inferior to other deep learning algorithms. The proposed algorithm effectively overcomes the challenge of slow training error rates, significantly enhancing the prediction rate and improving the actual prediction rate by 7%.

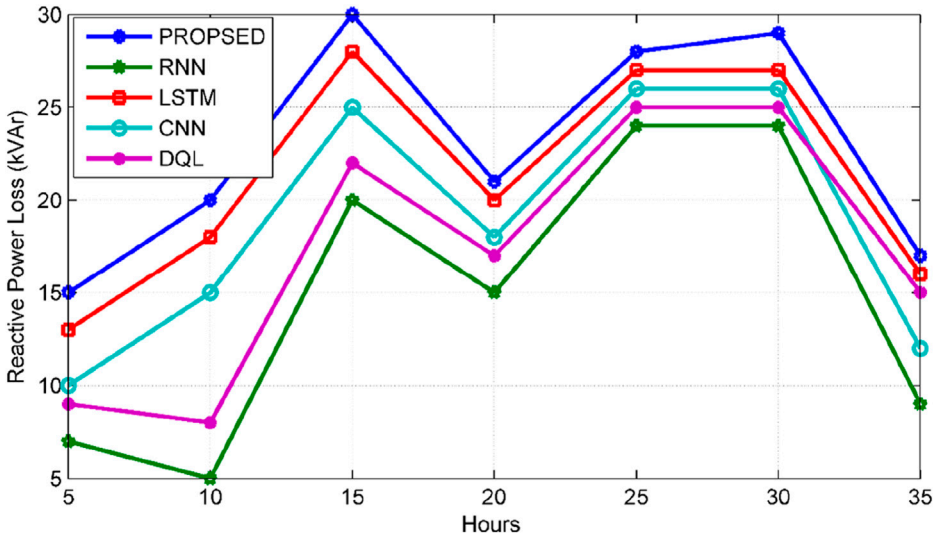


FIGURE 9
The predictive analysis of reactive power loss in the 69-bus distribution system.

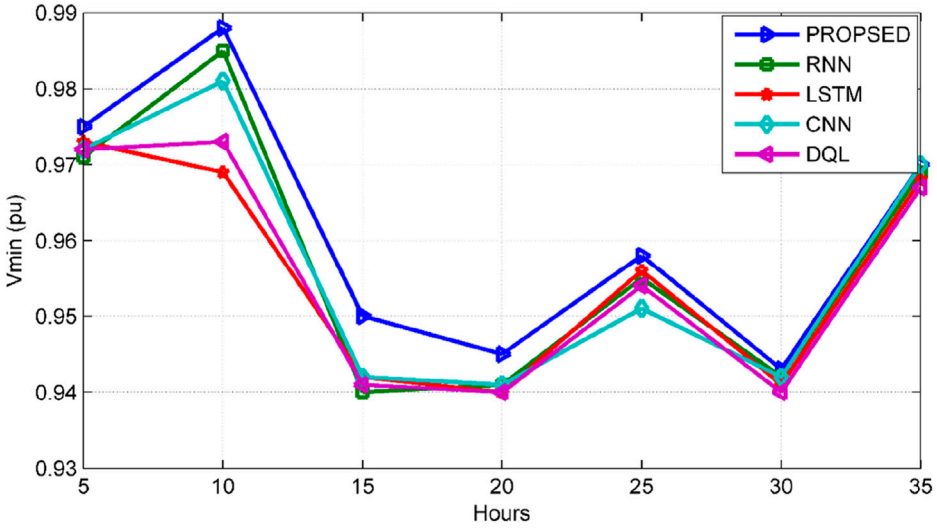


FIGURE 10
Minimum busbar voltage predictions for 33-bus distribution system.

3.5 Performance evaluation metrics

The assessment of the cascaded machine learning algorithm's performance is conducted using MATLAB software, comparing it with existing algorithms. This evaluation focuses on DG sizing placement, minimum busbar voltage, active power loss, and reactive power loss. The evaluation metrics employed to measure the algorithm's performance include R-squared/Coefficient of Determination and Mean Absolute Percentage Error (MAPE).

The objective is to assess the effectiveness of the cascaded machine learning algorithm in forecasting DG sizing placement, minimum busbar voltage, as well as both active and reactive power losses. This assessment involves comparing the algorithm's performance with established methods, utilizing R-squared and MAPE as crucial metrics.

This structured approach highlights the specific focus on evaluating and comparing the strength of the cascaded machine learning algorithm across several parameters against the existing algorithms.

3.5.1 R-squared/coefficient of determination

The evaluation of the model's performance and reliability involves calculating R-squared error metrics. In this evaluation, real values are utilized to confirm the precision of the model. R-squared (R2), known as the coefficient of determination, acts as a statistical indicator in regression models. It denotes the portion

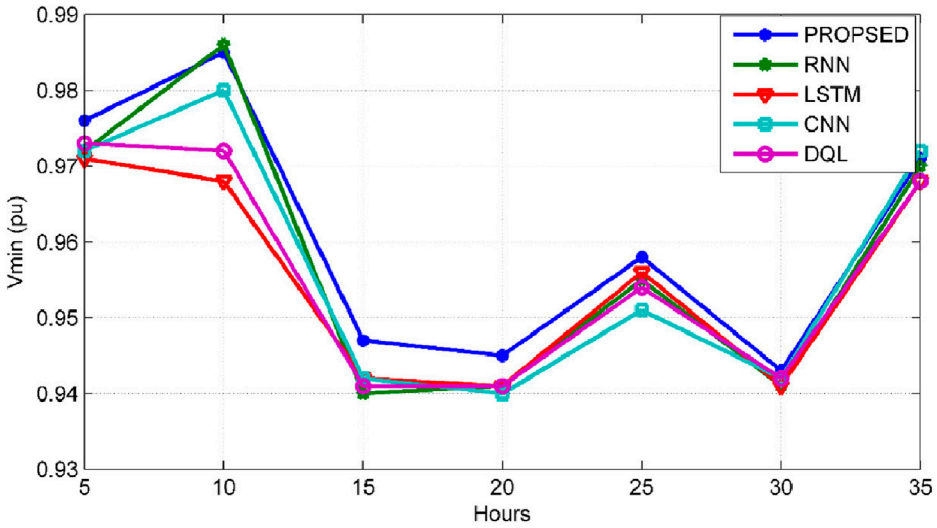


FIGURE 11
Minimum busbar voltage predictions for 69-bus distribution system.

of the variability in the reliant variable clarified by the independent one. R2 gauges how well the data fits the regression model, ranging between 0 and 1.

3.5.2 Mean Absolute Percentage Error (MAPE)

The assessment of forecast accuracy and outcome calculation is performed employing the Mean Absolute Percentage Error (MAPE). MAPE measures the correlation among absolute prediction inaccuracies and the real values. These metrics quantifies the accuracy of predictions relative to the actual data.

Tables 1–4 offer a comprehensive analysis of DG sizing, active power loss, reactive power loss, and minimum voltage estimation.

The results from the proposed model suggest that proposed algorithm surpasses CNN, RNN, LSTM, and DQL by 5.7%, 6.6%, 13.2%, and 14.2%, respectively, in estimating DG sizing within the 33-bus system, thereby enhancing the model's efficiency. Similarly, in the 69-bus system, the proposed model predicts DG sizing better by 5.6%, 9.3%, 15.8%, and 18.7% for CNN, RNN, LSTM, and DQL, respectively.

Comparing the proposed method with other state-of-the-art methods for active power loss estimation reveals significant performance improvements: 9.1% for CNN, 16.7% for RNN, 24.2% for LSTM, and 28.8% for DQL in the 33-bus test system. Likewise, in the 69-bus system, the proposed algorithm outperforms by 4.6%, 6.1%, 10.7%, and 23% for CNN, RNN, LSTM, and DQL, respectively.

Analyzing reactive power using the proposed method indicates superior performance over RNN, LSTM, DQL, and CNN with 5.3%, 10.5%, 13.2%, and 23.7%, respectively for the 33-bus test system. Similarly, in the 69-bus system, improvements over LSTM, CNN, DQL, and RNN are 3.6%, 7.1%, 10.7%, and 14.3%, respectively.

Predictions for Minimum Busbar voltage in the 33-bus system using LSTM, RNN, DQL, and CNN are less accurate compared to the proposed method by 0.2%, 0.3%, 0.4%, and 0.8%, respectively. Similarly, estimations for the 69-bus test system with the proposed method exhibit superiority by 0.1%, 0.3%, 0.5%, and 0.8%, respectively.

This section begins by analyzing the data relevant to the issue at hand. It then proceeds to compare the results obtained from the newly introduced models with those of the previously utilized ones. The efficacy of the proposed algorithm surpasses that of CNN, RNN, LSTM, and DQL across both the 33-bus and 69-bus systems. The utilization of the proposed algorithm for parameter estimation showcases superior effectiveness compared to relying on CNN, RNN, LSTM, and DQL. MAPE values indicate that the proposed algorithm consistently outperforms other algorithms for all parameters. R2 values exceeding 0.9 for various parameters underscore the effectiveness of the developed model. These findings affirm that the newly proposed model exhibits enhanced performance and can be relied upon to effectively address the challenge of predicting distribution system.

4 Conclusion

The study proposed the utilization of a cascaded machine learning algorithm for distributed generation (DG) allocation within distribution systems. It found that machine learning (ML) delivers impressive results for estimations based on single inputs. The CNN model exhibited significant effectiveness in forecasting different parameters like distributed generation (DG) sizing, system inefficiencies, and minimum voltage levels, especially in systems before incorporating the integration of distributed generation (DG).

The study aimed to assess the electrical distribution system by analyzing diverse factors, such as distributed generation (DG) power injection, active and reactive power losses, and minimum voltage, across both test systems, namely, IEEE 33-bus and IEEE 69-bus. Various machine learning (ML) models, such as CNN, RNN, LSTM, DQL, and Cascaded ML algorithm-based models, were developed and assessed using MAPE and R2 values. Notably, the Cascaded ML model exhibited high R2 values close to 1, indicating its suitability for accurate estimations. Validation with test data confirmed the effectiveness of these models. The findings emphasize the

effectiveness of the proposed method in precisely assessing the DG dimension and its consequential impacts on the power distribution network.

5 Future scope

To further enhance forecasting capabilities, the approach can be expanded by integrating additional features like line current or voltage drops during the model's development. These additional factors can be incorporated as extra inputs (features) in the input-output design to improve predictive accuracy. Moreover, the Machine Learning codes created in this study are versatile and can be adapted for other methods by adjusting the input data and utilizing the modified data to train different Machine Learning models.

However, in cases where there is a nonlinear correlation between input and output data, this model may not be appropriate for estimating DG integration. It is most effective for single-input predictions. Additionally, the suggested algorithm can be utilized to attain the most accurately predicted outcomes, including the distributed generation unit size, system losses, and minimum voltages in systems prior to the integration of DG.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/VasuBalan/RA.

Author contributions

AJ: Writing–review and editing, Writing–original draft, Validation, Software, Resources, Methodology, Investigation, Formal Analysis, Data curation, Conceptualization. SG: Supervision, Formal analysis, Validation, Visualization, Writing–original draft, 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.

References

Admasie, S., Basit, S., Bukhari, A., Gush, T., Haider, R., and Kim, C. H. (2020). Intelligent islanding detection of multi-distributed generation using artificial neural network based on intrinsic mode function feature. *J. Mod. Power Syst. Clean Energy* 8, 511–520. doi:10.35833/mpce.2019.000255

Agajie, T., Ferede, B., Khan, H., Alhelou, O. P. H., and Mahela, (2020). Optimal expansion planning of distribution system using grid-based multi-objective harmony search algorithm. *Comput. Electr. Eng.* 87, 106823. doi:10.1016/j.compeleceng.2020.106823

Alhussein, M., Aurangzeb, K., and Haider, S. I. (2020). Hybrid CNN-LSTM model for short-term individual household load forecasting. *IEEE Access* 8, 180544–180557. doi:10.1109/ACCESS.2020.3028281

Ali, Z. M., Diaaeldin, A. I. M., El-Rafei, H. M., HasanienAleem, A. Y., and Abdelaziz, (2021). A novel distributed generation planning algorithm via graphically-based network reconfiguration and soft open points placement using Archimedes optimization algorithm. *Ain Shams Eng. J.* 12, 1923–1941. doi:10.1016/j.asej.2020.12.006

Alnabi, L. A., Abbas, K., Dhaher, M. B., and Essa, (2022). Optimal allocation of distributed generation with reconfiguration by genetic algorithm using both Newton raphson and gauss seidel methods for power losses minimizing. *Int. J. Intelligent Eng. Syst.* 15, 464. doi:10.22266/ijies2022.0228.42

Arif, S., Muhammad, A., Hussain, T., Lie, T., Ahsan, H. A. S. M., and Khan, (2020). Analytical hybrid particle swarm optimization algorithm for optimal siting and sizing of distributed generation in smart grid. *J. Mod. Power Syst. Clean Energy* 8, 1221–1230. doi:10.35833/mpce.2019.000143

Bajaj, M., and Singh, A. K. (2021). Hosting capacity enhancement of renewable-based distributed generation in harmonically polluted distribution systems using passive harmonic filtering. *Sustain. Energy Technol. Assessments* 44, 101030. doi:10.1016/j.seta.2021.101030

Bajaj, M., Singh, A. K., Alowaidi, M., Sharma, N. K., Sharma, S. K., and Mishra, S. (2020). Power quality assessment of distorted distribution networks incorporating renewable distributed generation systems based on the analytic hierarchy process. *IEEE Access* 8, 145 713–145 737. doi:10.1109/access.2020.3014288

Battapothula, G., Yammani, C., and Maheswarapu, S. (2019). Multi-objective optimal planning of FCSs and DGs in distribution system with future EV load enhancement. *IET Electr. Syst. Transp.* 9, 128–139. doi:10.1049/iet-est.2018.5066

Belbachir, N., Zellagui, M., Lasmari, A., El-Bayeh, C. Z., and Bekkouche, B. (2021). Optimal integration of photovoltaic distributed generation in electrical distribution network using hybrid modified PSO algorithms. *Indonesian J. Electr. Eng. Comput. Sci.* 24, 50–60. doi:10.11591/ijeecs.v24.i1.pp50-60

Bhadoriya, J., Singh, A. R., and Gupta, (2022). A novel transient search optimization for optimal allocation of multiple distributed generator in the radial electrical distribution network. *Int. J. Emerg. Electr. Power Syst.* 23, 23–45. doi:10.1515/ijeeps-2021-0001

Bhusal, N., Gautam, M., Shukla, R. M., Benidris, M., and Sengupta, S. (2022). Coordinated data falsification attack detection in the domain of distributed generation using deep learning. *Int. J. Electr. Power & Energy Syst.* 134, 107345. doi:10.1016/j.ijepes.2021.107345

Bo, H., Nie, Y., and Wang, J. (2020). Electric load forecasting use a novelty hybrid model on the basic of data preprocessing technique and multi-objective optimization algorithm. *IEEE Access* 8, 13858–13874. doi:10.1109/access.2020.2966641

Chege, S. N., Murage, D. K., and Kihato, P. K. (2019). Optimal placement of distributed generation and capacitors in radial distribution networks using hybrid evolution programming algorithm. *Eur. J. Adv. Eng. Technol.* 6, 19–31. doi:10.5281/zenodo.10671905

Dheeban, S. S., and Selvan, N. B. M. (2021). ANFIS-based power quality improvement by photovoltaic integrated UPQC at distribution system. *IETE J. Res.* 69, 1–19.

Essallah, S., Khedher, A., and Bouallegue, A. (2019). Integration of distributed generation in electrical grid: optimal placement and sizing under different load conditions. *Comput. Electr. Eng.* 79, 106461. doi:10.1016/j.compeleceng.2019.106461

Eyüboglu, O., and Gül, Ö. (2021). Optimal allocation of multiple distributed generations including uncertainties in distribution networks by k-means clustering and particle swarm optimization algorithms. *Renew. Energy Power Qual. J.* 19.

Farh, H., Mh, A. M., Al-Shaalan, A. M., Eltamaly, A. A., and Al-Shamma, (2020). A novel crow search algorithm auto-drive PSO for optimal allocation and sizing of renewable distributed generation. *IEEE Access* 8, 27807–27820. doi:10.1109/access.2020.2968462

Feng, C., Sun, M., and Zhang, J. (2020). Reinforced deterministic and probabilistic load forecasting via Q-learning dynamic model selection. *IEEE Trans. Smart Grid* 11, 1377–1386. doi:10.1109/TSG.2019.2937338

Gawusu, S., Mensah, R. A., and Das, O. (2022). Exploring distributed energy generation for sustainable development: a data mining approach. *J. Energy Storage* 48, 104018. doi:10.1016/j.est.2022.104018

Haider, W., Hassan, S. J. U., Mehdi, A., Hussain, A., Adjayeng, G. O. M., and Kim, C.-H. (2021). Voltage profile enhancement and loss minimization using optimal placement and sizing of distributed generation in reconfigured network. *Machines* 9, 20. doi:10.3390/machines9010020

Hassan, S., Abdurrahman, Y., Sun, Z., and Wang, (2020). Optimization techniques applied for optimal planning and integration of renewable energy sources based on distributed generation: recent trends. *Cogent Eng.* 7, 1766394. doi:10.1080/23311916.2020.1766394

Hong, Y., Zhou, Y., Li, Q., Xu, W., and Zheng, X. (2020). A deep learning method for short-term residential load forecasting in smart grid. *IEEE Access* 8, 55785–55797. doi:10.1109/ACCESS.2020.2981817

Hu, H., Wang, L., Peng, L., and Zeng, Y.-R. (2020). Effective energy consumption forecasting using enhanced bagged echo state network. *Energy* 193, 116778. doi:10.1016/j.energy.2019.116778

Ismail, M., Shaaban, M. F., Naidu, M., and Serpedin, E. (2020). Deep learning detection of electricity theft cyber-attacks in renewable distributed generation. *IEEE Trans. Smart Grid* 11, 3428–3437. doi:10.1109/tsg.2020.2973681

Khasanov, M., Kamel, S., Rahmann, C., Hany, M., Hasanien, A., and Al-Durra, (2021). Optimal distributed generation and battery energy storage units integration in distribution systems considering power generation uncertainty. *IET Generation, Transm. Distribution* 15, 3400–3422. doi:10.1049/gtd2.12230

Kushal, T. R., Billah, M. S., and Illindala, (2020). Decision support framework for resilience-oriented cost-effective distributed generation expansion in power systems. *IEEE Trans. Industry Appl.* 57, 1246–1254. doi:10.1109/tia.2020.3047595

Lakum, A., and Mahajan, V. (2019). Optimal placement and sizing of multiple active power filters in radial distribution system using grey wolf optimizer in presence of nonlinear distributed generation. *Electr. Power Syst. Res.* 173, 281–290. doi:10.1016/j.epsr.2019.04.001

Lakum, A., and Mahajan, V. (2021). A novel approach for optimal placement and sizing of active power filters in radial distribution system with nonlinear distributed generation using adaptive grey wolf optimizer. *Eng. Sci. Technol. Int. J.* 24, 911–924. doi:10.1016/j.jestch.2021.01.011

Li, B., Zhang, J., He, Y., and Wang, Y. (2017). Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test. *IEEE Access* 5, 16324–16331. doi:10.1109/access.2017.2738029

Liu, W., Luo, F., Liu, Y., and Ding, W. (2019). Optimal siting and sizing of distributed generation based on improved nondominated sorting genetic algorithm II. *Processes* 7, 955. doi:10.3390/pr7120955

Lotfi, H. (2022). Optimal sizing of distributed generation units and shunt capacitors in the distribution system considering uncertainty resources by the modified evolutionary algorithm. *J. Ambient Intell. Humaniz. Comput.* 13, 4739–4758. doi:10.1007/s12652-021-03194-w

Menke, N. J., Hendrik, M., and Bornhorst, B. (2019). Distribution system monitoring for smart power grids with distributed generation using artificial neural networks. *Int. J. Electr. Power & Energy Syst.* 113, 472–480. doi:10.1016/j.ijepes.2019.05.057

Mo, H., and Sansavini, G. (2019). Impact of aging and performance degradation on the operational costs of distributed generation systems. *Renew. energy* 143, 426–439. doi:10.1016/j.renene.2019.04.111

Mohammadpourfard, M., Weng, Y., and Tajdinian, M. (2019). Benchmark of machine learning algorithms on capturing future distribution network anomalies. *IET Generation, Transm. Distribution* 13, 1441–1455. doi:10.1049/iet-gtd.2018.6801

Morales, J., Orduña, E., Villarroel, H., and Quispe, J. C. (2020). High-speed directional protection without voltage sensors for distribution feeders with distributed generation integration based on the correlation of signals and machine learning. *Electr. Power Syst. Res.* 184, 106295. doi:10.1016/j.epsr.2020.106295

Naguib, M., Mina, W. A., Omran, H., and Talaat, (2021). Performance enhancement of distribution systems via distribution network recon figuration and distributed generator allocation considering uncertain environment. *J. Mod. Power Syst. Clean Energy* 10, 647–655. doi:10.35833/mpce.2020.000333

Nsaif, Y. M., Lipu, A. M. H., Ayob, Y., Yusof, A., and Hussain, (2021). Fault detection and protection schemes for distributed generation integrated to distribution network: challenges and suggestions. *IEEE Access* 9, 142693–142717. doi:10.1109/access.2021.3121087

Ogunsina, A., Amos, M., Petinrin, O. O., Petinrin, E. N. O., Offornedo, J., Petinrin, G. O. O., et al. (2021). Optimal distributed generation location and sizing for loss minimization and voltage profile optimization using ant colony algorithm. *SN Appl. Sci.* 3, 248. doi:10.1007/s42452-021-04226-y

Onlam, A., Yodphet, D., Chatthaworn, R., Surawanitkun, C., Siritaratiwat, A., and Khunkitti, P. (2019). Power loss minimization and voltage stability improvement in electrical distribution system via network re configuration and distributed generation placement using novel adaptive shuffled frogs leaping algorithm. *Energies* 12, 553. doi:10.3390/en12030553

Panapakidis, I. P., Skiadopoulos, N., and Christoforidis, G. C. (2020). Combined forecasting system for short-term bus load forecasting based on clustering and neural networks. *IET Generation, Transm. Distribution* 14, 3652–3664. doi:10.1049/iet-gtd.2019.1057

Pereira, B. R., Costa, G. R. M. D., Contreras, J., and Mantovani, J. R. S. (2016). Optimal distributed generation and reactive power allocation in electrical distribution systems. *IEEE Trans. Sustain. Energy* 7, 975–984. doi:10.1109/tste.2015.2512819

Pham, T., Dinh Nguyen, B. T. T., and Dinh, H. (2021). Find optimal capacity and location of distributed generation units in radial distribution networks by using enhanced coyote optimization algorithm. *Neural Comput. Appl.* 33, 4343–4371. doi:10.1007/s00521-020-05239-1

Pham, T., Dinh Nguyen, L. C. T. T., and Kien, (2022). Optimal placement of photovoltaic distributed generation units in radial unbalanced distribution systems using MATLAB and OpenDSS-based cosimulation and a proposed metaheuristic algorithm. *Int. Trans. Electr. Energy Syst.* 2022, 1–21. doi:10.1155/2022/1446479

Purlu, M., and Turkay, B. E. (2021). Estimating the distributed generation unit sizing and its effects on the distribution system by using machine learning methods. *Elektron. Elektrotech.* 27, 24–32. doi:10.5755/j02.eie.28864

Qiao, W., Lu, H., Zhou, G., Azimi, M., Yang, Q., and Tian, W. (2020). A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer. *J. Clean. Prod.* 244, 118612. doi:10.1016/j.jclepro.2019.118612

Rathore, A., and Patidar, N. P. (2021). Optimal sizing and allocation of renewable based distribution generation with gravity energy storage considering stochastic nature using particle swarm optimization in radial distribution network. *J. Energy Storage* 35, 102282. doi:10.1016/j.est.2021.102282

Routray, A., Mistry, K. D., Arya, S. R., and Chittibabu, B. (2021). "Applied machine learning in wind speed prediction and loss minimization in unbalanced radial distribution system". in *Energy sources, Part A: recovery, utilization, and environmental effects*, 1–21.

Sambaiah, K., Sampangi, T., and Jayabarathi, (2019). Optimal allocation of renewable distributed generation and capacitor banks in distribution systems using salp swarm algorithm. *Int. J. Renew. energy Res.* 9, 96–107. https://doi.org/10.20508/ijrer.v9i1.8581.g75

Selim, A., Kamel, S., Mohamed, A. A., and Elattar, E. E. (2021). Optimal allocation of multiple types of distributed generations in radial distribution systems using a hybrid technique. *Sustainability* 13, 6644. doi:10.3390/su13126644

Shaheen, M. A., Hany, M., Hasanien, S. F., Mekhamer, H., and Talaat, (2019). Optimal power flow of power systems including distributed generation units using sunflower optimization algorithm. *IEEE Access* 7, 109289–109300. doi:10.1109/access.2019.2933489

Tran, T., The, K., Truong, D. N. H., and Vo, (2020). Stochastic fractal search algorithm for reconfiguration of distribution networks with distributed generations. *Ain Shams Eng. J.* 11, 389–407. doi:10.1016/j.asej.2019.08.015

Vai, V., Suk, S., Lorm, R., Chhlonh, C., Eng, S., and Bun, L. (2021). Optimal reconfiguration in distribution systems with distributed generations based on modified sequential switch opening and exchange. *Appl. Sci.* 11, 2146. doi:10.3390/app11052146

Venkatesan, C., Kannadasan, R., Alsharif, M. H., Kim, M.-K., and Nebhen, J. (2021). A novel multiobjective hybrid technique for siting and sizing of distributed generation and capacitor banks in radial distribution systems. *Sustainability* 13, 3308. doi:10.3390/su13063308

Wang, H., Lei, Z., Zhang, X., Zhou, B., and Peng, J. (2019). A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* 198, 111799. doi:10.1016/j.enconman.2019.111799

Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., et al. (2020). Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Convers. Manag.* 214, 112909. doi:10.1016/j.enconman.2020.112909

Wu, D., Wang, B., Precup, D., and Boulet, B. (2020). Multiple kernel learning-based transfer regression for electric load forecasting. *IEEE Trans. Smart Grid* 11, 1183–1192. doi:10.1109/TSG.2019.2933413

Yang, Y., Wang, Z., Gao, Y., Wu, J., Zhao, S., and Ding, Z. (2022). An effective dimensionality reduction approach for short-term load forecasting. *Electr. Power Syst. Res.* 210, 108150. doi:10.1016/j.epsr.2022.108150

Yang, Y., Wang, Z., Zhao, S., and Wu, J. (2023). An integrated federated learning algorithm for short-term load forecasting. *Electr. Power Syst. Res.* 214, 108830. doi:10.1016/j.epsr.2022.108830

Yılmaz, A., Küçüker, A., and Bayrak, G. (2022). Automated classification of power quality disturbances in a SOFC&PV-based distributed generator using a hybrid machine learning method with high noise immunity. *Int. J. Hydrogen Energy* 47, 19797–19809. doi:10.1016/j.ijhydene.2022.02.033

Yin, Z., Tu, J., and Xu, Y. (2019). Development of a kernel extreme learning machine model for capacity selection of distributed generation considering the characteristics of electric vehicles. *Appl. Sci.* 9, 2401. doi:10.3390/app9122401

Zhang, Y., Wang, X., and Wang, J. (2020). Deep reinforcement learning based volt-var optimization in smart distribution systems. *IEEE Trans. Smart Grid* 12, 361–371. doi:10.1109/tsg.2020.3010130

Zhao, H., Zhang, H., Ge, Y., Gao, W., Li, J., Su, B., et al. (2019). "Reliability analysis of distribution network based on the deep coupling of high penetration distributed generation and energy storage," in *IEEE sustainable power and energy conference* (Beijing, China: iSPEC), 2561–2565.

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 10 June 2024
DOI 10.3389/fenrg.2024.1409957

Check for updates

OPEN ACCESS

EDITED BY
Yang Yu,
Nanjing University of Posts and Telecommunications, China

REVIEWED BY
Kenneth E. Okedu,
Melbourne Institute of Technology, Australia
Bowen Zhou,
Northeastern University, China

*CORRESPONDENCE
Zhenning Huang,
aaa1748556513@163.com

RECEIVED 31 March 2024
ACCEPTED 22 May 2024
PUBLISHED 10 June 2024

CITATION
Huang Z, Sun N, Shao H and Li Y (2024), Ultra-short-term prediction of microgrid source load power considering weather characteristics and multivariate correlation.
Front. Energy Res. 12:1409957.
doi: 10.3389/fenrg.2024.1409957

COPYRIGHT
© 2024 Huang, Sun, Shao 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.

Ultra-short-term prediction of microgrid source load power considering weather characteristics and multivariate correlation

Zhenning Huang[1]*, Ning Sun[2], Huaqiang Shao[2] and Yunjing Li[2]

[1]State Grid Shandong Electric Power Company, Shandong, Jinan, China, [2]State Grid Yantai Power Company, Shandong, Jinan, China

Multiple microgrids interconnect to form a microgrid cluster. To fully exploit the comprehensive benefits of the microgrid cluster, it is imperative to optimize dispatch based on the matching degree between the sources and loads of each microgrid. The power of distributed energy sources such as wind and photovoltaic systems and the sensitive loads in microgrids is related to the regional weather characteristics. Given the relatively small geographical scope of microgrid areas and the fact that distributed energy sources and loads within the grid share the same weather characteristics, simultaneous ultra-short-term forecasting of power for both sources and loads is essential in the same environmental context. Firstly, the introduction of the multi-variable uniform information coefficient (MV-UIC) is proposed for extracting the correlation between weather characteristics and the sequences of source and load power. Subsequently, the application of factor analysis (FA) is introduced to reduce the dimensionality of input feature variables. Drawing inspiration from the concept of combination forecasting models, a combined forecasting model based on Error Back Propagation Training (BP), Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory Neural Network (BiLSTM) is constructed. This model is established on the MV-UIC-FA foundation for the joint ultra-short-term forecasting of source and load power in microgrids. Simulation is conducted using the DTU 7K 47-bus system as an example to analyze the accuracy, applicability, and effectiveness of the proposed joint forecasting method for sources and loads.

KEYWORDS

weather characteristics, multivariable unified information coefficient, source and load power, joint prediction, machine learning

1 Introduction

In microgrids, there exists a substantial presence of distributed generation (DG) sources such as wind and photovoltaic, along with actively time-varying sensitive loads. The power output of DG and load power are both influenced by complex factors such as regional weather, date, and special events (Zhu et al., 2023). As microgrid deployment and utilization expand, neighboring microgrids interconnect to form coexistence of distributed generation and loads in the same environment, it is imperative to simultaneously conduct joint prediction of source and load power under the same weather characteristics (Yu et al., 2024).

To fully exploit the comprehensive benefits of microgrid clusters, it is necessary to coordinate and optimize the operation within each microgrid cluster and between microgrid clusters and distribution networks based on the matching degree of sources and loads in each microgrid. Ultra-short-term prediction of source and load power serves as the foundation for this endeavor. Given the relatively small geographical area where microgrids are located and the coexistence of distributed generation and loads in the same environment, it is imperative to simultaneously conduct joint prediction of source and load power under the same weather characteristics (Wang et al., 2024).

Microgrid source and load power ultra-short-term prediction methods encompass mathematical statistical approaches (Safari et al., 2018) and artificial intelligence methods (Zhu et al., 2023). Artificial intelligence methods excel in capturing the nonlinear relationship between inputs and outputs, demonstrating robust data analysis and forecasting capabilities. They have emerged as pivotal techniques for ultra-short-term prediction of microgrid source and load power (Zhang et al., 2024). Internationally recognized experts and scholars, considering factors such as meteorological and calendar features, have employed various single prediction models like Error Back Propagation Training (BP), deep recurrent neural network (DRNN), and Long Short-Term Memory (LSTM) to forecast short-term power for microgrid DG and loads. In response to the escalating electricity demand and restructuring of power systems, researchers have proposed a long-term electricity demand forecasting method based on BP (Masoumi et al., 2020). This approach utilizes a Time Series Neural Network (TSNN) structure, employing forward propagation of input load data, error computation, and weight updating through backpropagation in training steps, thereby achieving self-learning and self-organization. To enhance photovoltaic (PV) generation prediction accuracy, researchers have developed a forecasting algorithm based on LSTM (Hossain and Mahmood, 2020). This algorithm combines synthesized weather forecasts with historical solar radiation data and publicly available sky type predictions for the host city, utilizing the K-means algorithm for dynamic sky type classification. This approach significantly improves prediction accuracy, achieving an increase in accuracy ranging from 33% to 44.6% compared to predictions using fixed sky types. In another study, G. W. Chang and H.J. Lu integrated grey data preprocessors with DRNN for day-ahead output prediction in photovoltaic generation (Chang and Lu, 2018). However, due to the distinct advantages of different prediction models, achieving optimal predictive performance with a single model often proves challenging (Zhu et al., 2019). To address this, experts and scholars adopt a "modal decomposition-combined prediction" approach: firstly, utilizing modal decomposition methods to break down historical data of sources and loads into components with different frequencies, thereby reducing the complexity of input data; secondly, introducing a combined model approach, selecting prediction models with varying performances for different frequency modal components, and ultimately obtaining the final prediction results through summation and reconstruction. For instance, adaptive noise-aided complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) was employed to decompose the raw data of building sub-item energy generation (Lin, 2022). Subsequently, predictions for different modal components were conducted using BiLSTM. The final photovoltaic power generation was obtained through summation and reconstruction. Another study combined three models, optimized extreme learning machine (ELM), backpropagation neural network (BPNN), and dynamic recursive neural network (ELMAN), for short-term forecasting of wind power plant output (Ma et al., 2023). In a separate study, variational mode decomposition (VMD) was utilized to decompose load raw data into different frequency modal components (Yue et al., 2023). Subsequently, a Bagging ensemble ultra-short-term multivariate load forecasting method was developed based on gated recurrent unit (GRU), LSTM, and BiLSTM models, leading to enhanced prediction accuracy. In the field of microgrid power forecasting, predictions of source power and load power are typically conducted independently based on their respective environmental factors. However, this approach may overlook an important reality: under the same environmental conditions, weather characteristics may simultaneously affect both the source side (such as photovoltaic and wind power) and the load side of the microgrid. For instance, sunny weather may increase the output of photovoltaic generation while also raising the electricity demand for cooling devices like air conditioners. Therefore, decoupling the predictions of source power and load power may lead to reduced accuracy in forecasting results, impacting the economic dispatch and stable operation of the microgrid. Currently, research on how to comprehensively consider the simultaneous impact of weather characteristics on both the source and load sides of microgrids is relatively limited. Most existing models focus on predicting power for a single energy source type or only consider the influence of weather factors on load demand. Such independently predictive methods may fail to fully capture the comprehensive effects of weather changes on the overall performance of microgrids. To enhance the accuracy and reliability of microgrid power forecasting, future research needs to develop more comprehensive models that can simultaneously consider the generation characteristics of multiple energy sources and the response of load demand to weather changes.

To predict both the source and load power in a microgrid under the same weather conditions simultaneously, it is necessary to analyze the concurrent correlation between the two variables and the weather features. Extracting weather characteristics that have a significant impact on the source and load power in the microgrid enhances predictive capability. Common methods for feature extraction include the Pearson coefficient (Xu et al., 2023), Spearman coefficient (Qun et al., 2023), Maximum Information Coefficient (MIC) (Reshef et al., 2011), and Uniform Information Coefficient (UIC) (Jiang et al., 2023). Reference Jiang et al. (2023) utilized the Uniform Information Coefficient (UIC) to analyze the correlation between weather characteristics and load power. Compared to the other three algorithms, UIC is specifically designed for analyzing relationships among multidimensional variables, making it more suitable for handling complex meteorological data and load power data, which are often multidimensional. Moreover, it is computationally more efficient. It achieves this by employing a simplified technique based on uniformly partitioned data grids, replacing the dynamic programming steps in MIC computation, thus reducing computational costs. The aforementioned approaches focus solely on extracting the correlation between a single dependent variable

(e.g., DG output or load power) and a single independent variable (weather feature), thereby unable to capture the correlation between multiple dependent variables (DG output and load power) and a single independent variable (weather feature). Addressing this limitation, this study investigates the simultaneous correlation between source and load power in a microgrid and weather features, conducting research on the joint ultra-short-term prediction of source and load power in a microgrid. Additionally, commonly used dimensionality reduction algorithms include Principal Component Analysis (PCA) (Wang et al., 2023), Independent Component Analysis (ICA) (Kobayashi and Iwai, 2018), Factor Analysis (FA) (Ramirez et al., 2019; Wu et al., 2024), etc. FA merges numerous features into several representative common factors to extract latent factors among features, accurately capturing the relevant information in the data (Zhou et al., 2020). FA is particularly effective in capturing the underlying structure of data by reducing the dimensionality and identifying the shared variance among variables. It helps uncover the latent factors that explain the correlations and patterns within the dataset, facilitating a deeper understanding of the relationships among the features.

In conclusion, this manuscript contemplates the impact of weather features in the region of a microgrid on DG and load power simultaneously. A joint ultra-short-term prediction model for source and load power in a microgrid is proposed. Initially, the concept of the Multi-Variable Uniform Information Coefficient (MV-UIC) is introduced to analyze and compute the correlation coefficients between weather features and the sequences of microgrid source and load power, facilitating the elimination of redundant features. To diminish the dimensionality of input features for the prediction models of source and load power, FA is employed. Addressing the pronounced nonlinearity and non-stationarity in microgrid source and load power, a model is established by amalgamating BP, LSTM, and BiLSTM models, considering the correlation between weather features and multiple variables. Using the DTU 7K 47-bus as an example in a real system (Baviskar et al., 2021), specifically with 3 wind farms serving as DG and an aggregated load, the proposed prediction model is pre-trained using historical data from the power grid dataset. The accuracy, applicability, and effectiveness of the proposed joint prediction method for source and load are then analyzed in the DTU 7K 47-bus system.

2 A weather feature extraction method based on MV-UIC

The Uniform Information Coefficient (UIC) algorithm, pioneered by Mousavi and Baranuk (2022), introduces an innovative methodology for feature extraction. UIC facilitates the analysis of the correlation between two univariate variables, making it particularly well-suited for addressing feature extraction challenges in large-scale datasets. Let $\mathbf{A} = [a_1, \ldots, a_n]$ and $\boldsymbol{B} = [b_1, \ldots, b_n]$ denote two sets of feature vectors with a sequence length of n. The model for the UIC is shown in Eq. 1.

$$I_{\mathrm{UIC}}(\mathbf{A};\mathbf{B}) = \frac{I(\mathbf{A};\mathbf{B})}{\log_2(\min\{r,s\})} \tag{1}$$

where $I_{\mathrm{UIC}}(\mathbf{A};\mathbf{B})$ represents the Uniform Information Coefficient between **A** and **B**; $I(\mathbf{A};\mathbf{B})$ denotes the Mutual Information Coefficient between **A** and **B**; r and s correspondingly indicate the segmentation numbers for **A** and **B**; $\min\{r, s\}$ represents the minimum value between r and s.

In order to ascertain the simultaneous correlation of the dependent variables, source and load power, with weather characteristics in a microgrid, this paper adopts an extension of the method presented in Wang (2020) and Ng et al. (2023), which expands from the maximum information coefficient between two variables to the multivariate case. The UIC for the two variables is extended, introducing the Multi-variable Uniform Information Coefficient (MV-UIC) algorithm. The specific definition is outlined as follows: The dataset D consists of three variables: (P_{DG}, P_{L}), and H. H represents the independent variable, denoting the regional weather feature vector, while (P_{DG}, P_{L}) signify the source and load power vectors, respectively, in a microgrid. H is allocated to the x-axis as $X = [x_i]$, where "n" denotes the sequence length. Similarly, (P_{DG}, P_{L}) are allocated to the y-axis as $Y = [y_i]$. H is uniformly divided into r blocks, and (P_{DG}, P_{L}) into s blocks. After partitioning dataset D, $r \times s$ grids are obtained, with each grid representing a subset of data points. The ratio of the number of points falling into corresponding grid to the total number of points defines the approximate probability density of that grid. Subsequently, the mutual information coefficient between the weather feature variable H and (P_{DG}, P_{L}) is derived. Normalization is then performed to obtain the multivariate unified information coefficient between the two variables P_{DG}, P_{L}, and the univariate H in dataset D. The steps for obtaining MV-UIC are as follows:

(1) For a given three-variable dataset D and positive integers r, s, where $r, s \geq 2$, $D = \{(P_{\mathrm{DG}}, P_{\mathrm{L}}), H\}$, the mutual information between them, denoted as $I_{\mathrm{MI}}(D, r, s, (P_{\mathrm{DG}}, P_{\mathrm{L}}); H)$:

$$I_{MI}(D, r, s, (P_{DG}, P_L), H) = \sum_{x=1}^{r}\sum_{y=1}^{s} p(xy)\log_2\frac{p(xy)}{p(x)p(y)} \tag{2}$$

where $p(x)$ represents the edge probability density of the regional weather characteristic variable H uniformly divided into r grids, $p(y)$ represents the edge probability density of the source and load power vectors (P_{DG}, P_{L}) uniformly divided into s grids, and $p(xy)$ represents the joint probability density of the dataset D divided into $r \times s$ grids.

According to the uniform division method, the length of each segment after evenly dividing the x and y axes into r and s is:

$$\begin{cases} d_x = \dfrac{x_{\max} - x_{\min}}{r}, 2 \leq r \leq 1 + \dfrac{n^{0.6}}{2} \\ d_y = \dfrac{y_{\max} - y_{\min}}{s}, 2 \leq s \leq 1 + \dfrac{n^{0.6}}{2} \end{cases} \tag{3}$$

where d_x and d_y represent the lengths of partition units for X and Y, respectively; x_{max} and x_{min} denote the maximum and minimum values of the feature vector X; y_{max} and y_{min} correspondingly signify the maximum and minimum values of the feature vector Y; r and s designate the number of segments for X and Y, respectively; and $n^{0.6}$ symbolizes the size of the partition grid, typically chosen as the 0.6th power of the data volume (Wang, 2020).

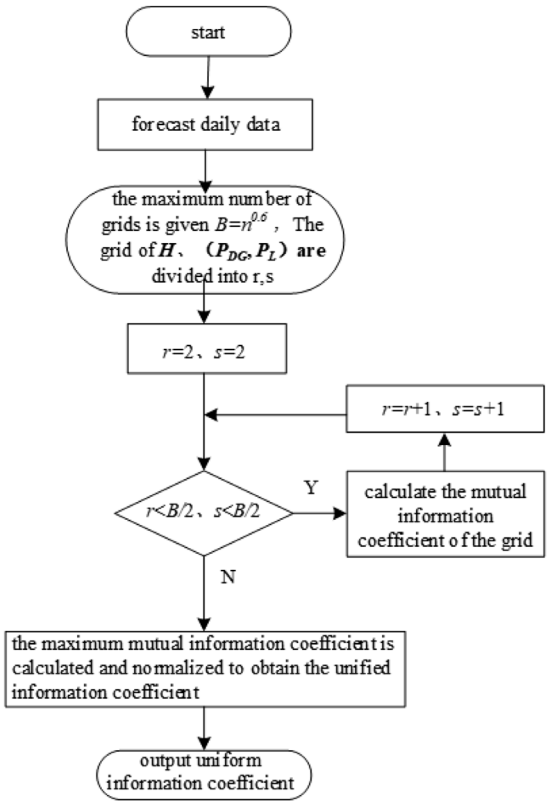


FIGURE 1
Flow chart of the MV-UIC algorithm.

(2) The formula for calculating the MV-UIC between the regional weather feature vector H and the source and load power vectors (P_{DG}, P_L) in a microgrid is as follows:

$$\mathrm{MV-MIC}\{(P_{DG}, P_L); H\} = \max_{r,s<\partial} \frac{I_{MI}(D, r, s, (P_{DG}, P_L), H)}{\log_2(\min\{r, s\})} \quad (4)$$

where $I(D, r, s, (P_{DG}, P_L), H)$ denotes the mutual information among the three entities; $\log_2(\min\{r, s\})$ represents normalization, with min $\{r, s\}$ being the minimum value between r and s, and ∂ denoting the maximum grid partition number ($r, s < \partial$). The flow chart of the MV-UIC algorithm is shown in Figure 1.

Based on Eqs 2–4, the MV-UIC between microgrid source, load power, and weather features can be computed. A higher value indicates a stronger correlation between the respective weather feature and the source/load power. Selecting weather features with high correlations as inputs for source/load power prediction models helps filter out remaining features, thereby mitigating issues related to excessive variables and redundant computations.

3 Inputfeature dimensionality reduction based on FA

Due to the strong interrelationships among weather feature variables, regression analysis encounters a certain degree of collinearity issue. FA serves as a multivariate statistical method that, by solving the correlation matrix of variables, identifies common factors describing relationships among numerous variables and simplifies data, thereby reducing the dimensionality of the dataset. The fundamental principles and computational procedures of FA are detailed in Wu et al. (2024), wherein the basic model entails the linear relationship between observed variables and common factors as Eq. 5:

$$\mathbf{\Psi} = \iota^{*}\mathbf{F} + \mathbf{E} \quad (5)$$

In the equation, $\mathbf{\Psi}$ represents the matrix of observed variables; $\mathbf{F}$ stands for the matrix of common factors; ι denotes the factor loading matrix, illustrating the relationship between each observed variable and the common factors; $\mathbf{E}$ signifies the matrix of factor variances.

Utilizing the MV-UIC obtained in Section 1, weather feature variables with similar attributes are grouped together and represented by a common factor. Analyzing the correlation between variables involves solving the eigenvalues and corresponding orthogonal eigenvectors of the correlation matrix. Based on the eigenvalues of the correlation matrix, the variance contribution rate and cumulative contribution rate of common factors are computed, with a cumulative contribution rate exceeding 85% serving as the criterion for determining the number of common factors. Subsequently, factor matrix rotation yields the factor loading matrix. Factor scores are then calculated using regression analysis. Higher values in the factor score matrix indicate a more significant representation of the feature by the respective factor in the dataset. Dimensionality reduction of input features for source/load power prediction is conducted based on factor scores.

4 A short term joint prediction model for microgrid source and load power considering weather characteristics and multivariable correlation

In microgrid systems, predicting source and load power is crucial for stable operation. Due to their nonlinearity and non-stationarity, single models struggle to capture these complexities, leading to poor performance. Empirical Mode Decomposition (EMD) decomposes source-load sequences into Intrinsic Mode Functions (IMFs), enhancing prediction accuracy by describing variations and periodicities. To improve predictive performance further, joint prediction methods integrate multiple models' advantages. Weighting different models appropriately creates a comprehensive model considering various characteristics and IMFs, yielding more accurate results. Additionally, predicting source and load power under similar weather conditions requires analyzing their correlation with weather features. Traditional methods fail to capture this correlation simultaneously, unlike the Multi-variable Uniform Information Coefficient (MV-UIC), which evaluates it effectively. MV-UIC's application enables feasible joint prediction of source and load, quantifying the correlation between multiple dependent variables and a single independent variable, aiding in constructing precise prediction models.

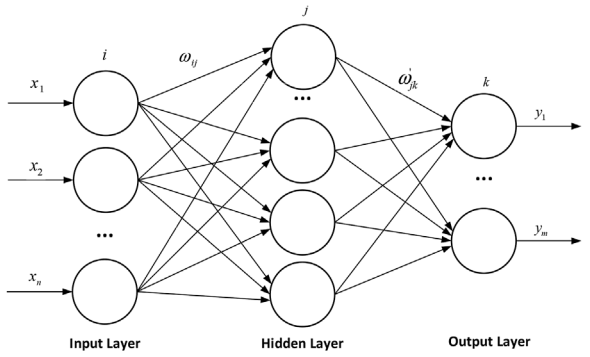


FIGURE 2
BP network with three layers.

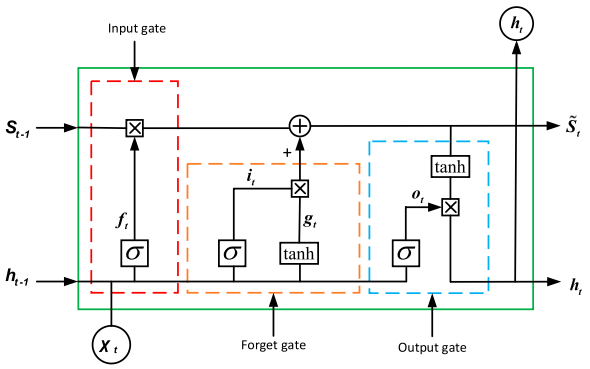


FIGURE 3
Basic unit of LSTM network.

4.1 BP network

The BP network is a multi-layer feedforward neural network. The topology of a three-layer BP network is illustrated in Figure 2, encompassing an input layer, an output layer, and a single hidden layer. Each neuron is connected to all neurons in the subsequent layer, with no interconnections among neurons within the same layer.

The BP network minimizes error using gradient descent. Standard BP lacks momentum consideration, causing slow convergence. Enhanced BP integrates momentum to reduce oscillations and hasten convergence. The objective function is defined accordingly. The objective function is defined as Eq. 6:

$$J_{\mathrm{BP}} = \frac{1}{2}\sum_{\delta=1}^{N_{\mathrm{BP}}}\sum_{\phi=1}^{m_{\mathrm{BP}}}\left(z_{\phi}^{\delta} - c_{\phi}^{\delta}\right)^2 \tag{6}$$

where c_{ϕ}^{δ} represents the output of node ϕ when sample δ is applied; z_{ϕ}^{δ} denotes the target value of output bus ϕ for sample δ, m_{BP} is the dimension of the output variable, and N_{BP} is the number of training samples.

4.2 LSTM

The LSTM represents an enhanced version of the Recurrent Neural Network (RNN). Introduced and subsequently refined with additional forget gates, the improved LSTM addresses the issue of "vanishing gradients" encountered during model training. Capable of learning both short-term and long-term dependencies in time series data, it stands as one of the most successful RNN architectures, finding applications across various domains. The fundamental unit of an LSTM network, as depicted in Figure 3.

The fundamental unit of an LSTM network comprises forget gates, input gates, and output gates. The forget gate determines the extent of memory to be retained from the state cell, influenced by the input $\boldsymbol{\chi}_t$, previous state $\tilde{\mathbf{S}}_{t-1}$, and the intermediary output $\mathbf{h}_{t-1}$. The input gate decides the vectors to be preserved within the state cell, with $\boldsymbol{\chi}_t$ undergoing transformations via sigmoid and tanh functions. The intermediary output $\mathbf{h}_{t-1}$ is jointly determined by the updated state $\tilde{\mathbf{S}}_t$ and the output $\mathbf{o}_t$, as outlined in Eqs 7–12:

$$\boldsymbol{f}_t = \sigma\left(\boldsymbol{W}_{f\chi}\boldsymbol{\chi}_t + \boldsymbol{W}_{fh}\boldsymbol{h}_{t-1} + \boldsymbol{\beta}_f\right) \tag{7}$$

$$\boldsymbol{i}_t = \sigma\left(\boldsymbol{W}_{i\chi}\boldsymbol{\chi}_t + \boldsymbol{W}_{ih}\boldsymbol{h}_{t-1} + \boldsymbol{\beta}_i\right) \tag{8}$$

$$\boldsymbol{g}_t = \text{ƛ}\left(\boldsymbol{W}_{g\chi}\boldsymbol{\chi}_t + \boldsymbol{W}_{gh}\boldsymbol{h}_{t-1} + \boldsymbol{\beta}_g\right) \tag{9}$$

$$\boldsymbol{o}_t = \sigma\left(\boldsymbol{W}_{o\chi}\boldsymbol{\chi}_t + \boldsymbol{W}_{oh}\boldsymbol{h}_{t-1} + \boldsymbol{\beta}_o\right) \tag{10}$$

$$\tilde{\boldsymbol{S}}_t = \boldsymbol{g}_t \odot \boldsymbol{i}_t + \tilde{\boldsymbol{S}}_{t-1} \odot \boldsymbol{f}_t \tag{11}$$

$$\boldsymbol{h}_t = \text{ƛ}\left(\tilde{\boldsymbol{S}}_t\right) \odot \boldsymbol{o}_t \tag{12}$$

where $\boldsymbol{f}_t$, $\boldsymbol{i}_t$, $\boldsymbol{g}_t$, $\boldsymbol{o}_t$, $\boldsymbol{h}_t$ and $\tilde{\boldsymbol{S}}_t$ represent the states of the forget gate, input gate, input node, output gate, intermediary output, and state unit, respectively. $\boldsymbol{W}_{f\chi}$, $\boldsymbol{W}_{fh}$, $\boldsymbol{W}_{i\chi}$, $\boldsymbol{W}_{ih}$, $\boldsymbol{W}_{g\chi}$, $\boldsymbol{W}_{gh}$, $\boldsymbol{W}_{o\chi}$, and $\boldsymbol{W}_{oh}$ denote the matrix weights for the respective gates multiplied by the input $\boldsymbol{\chi}_t$ and intermediary output $\boldsymbol{h}_{t-1}$. $\boldsymbol{\beta}_f$, $\boldsymbol{\beta}_i$, $\boldsymbol{\beta}_g$, and $\boldsymbol{\beta}_o$ are the bias terms for the corresponding gates. The symbol $\odot$ represents element-wise multiplication in vectors, σ denotes the sigmoid function transformation, and ƛ represents the tanh function transformation.

4.3 BiLSTM

The BiLSTM is an advanced enhancement of the conventional unidirectional LSTM, integrating both a forward LSTM layer and a backward LSTM layer, each influencing the output. While the unidirectional LSTM adeptly utilizes historical data to mitigate long-distance dependency issues, the BiLSTM benefits from the incorporation of both forward and backward sequence information, thoroughly considering past and future data to significantly enhance model prediction accuracy. The architecture of the BiLSTM, as illustrated in Figure 4.

The architecture of the BiLSTM involves updates to the hidden layers of the forward LSTM, the backward LSTM, and the process leading to the final output of the BiLSTM, delineated in Eqs 13–15:

$$\boldsymbol{\varsigma}_t = f_{\mathrm{BiLSTM},1}\left(\omega_1\boldsymbol{\chi}_t + \omega_2\boldsymbol{\varsigma}_{t-1}\right) \tag{13}$$

$$\boldsymbol{\xi}_t = f_{\mathrm{BiLSTM},2}\left(\omega_3\boldsymbol{\chi}_t + \omega_5\boldsymbol{\xi}_{t+1}\right) \tag{14}$$

$$\boldsymbol{\gamma}_t = f_{\mathrm{BiLSTM},3}\left(\omega_4\boldsymbol{\varsigma}_t + \omega_6\boldsymbol{\xi}_t\right) \tag{15}$$

where $f_{\mathrm{BiLSTM},1}$, $f_{\mathrm{BiLSTM},2}$ and $f_{\mathrm{BiLSTM},3}$ denote the activation functions between the different layers; $\boldsymbol{\varsigma}_t$ and $\boldsymbol{\xi}_t$ represent the corresponding LSTM hidden states for the forward and backward iterations, respectively; $\boldsymbol{\chi}_t$ represents the corresponding input data; $\boldsymbol{\gamma}_t$ represents the corresponding input data; $\omega_1 \ldots \omega_6$ represent the corresponding weight of each layer.

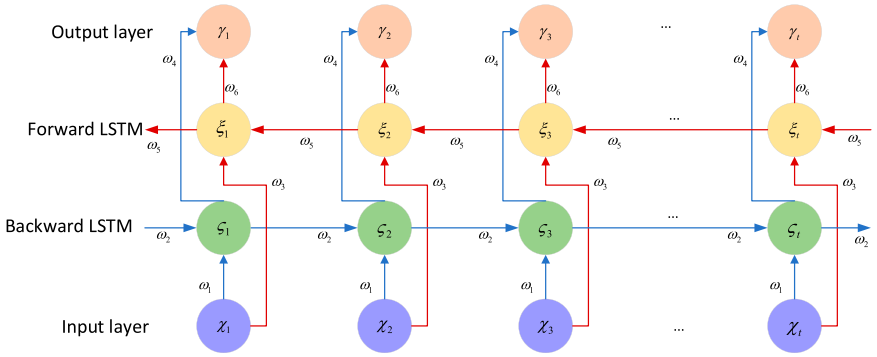


FIGURE 4
Structure of BiLSTM network.

4.4 VMD

The specific steps of the VMD algorithm are described as follows:

(1) Define the variational problem: In order to decompose the given original sequence $f_{\mathrm{VMD}}(\ell)$ into $\tilde{K}$ variational mode components u_κ with different central frequency bandwidths, and the sum of the estimated bandwidths of each mode is the minimum, the variational constraint expression is defined as Eq. 16:

$$\begin{cases} \min\limits_{\{u_\kappa\},\{\omega_\kappa\}} \left\{ \sum\limits_{\kappa=1}^{\tilde{K}} \left\| \partial_t \left[\vartheta_\kappa(\ell) \right] e^{-j\omega_\kappa \ell} \right\|_2^2 \right\} \\ s.t. \sum\limits_{\kappa=1}^{\tilde{K}} u_\kappa(\ell) = f_{\mathrm{VMD}}(\ell) \end{cases} \tag{16}$$

where the original sequence $f_{\mathrm{VMD}}(\ell)$ is the set of optimal similar daily load sequence; $\{u_\kappa\}\{\omega_\kappa\}$, which represents the set of all submodes and their corresponding center frequencies respectively; $\vartheta_\kappa(\ell)$ represents the analytical signal of the κ submode after being demodulated by Hilbert; $u_\kappa(\ell)$ represents the modal function of the κ submode.

(2) The formula for the Lagrangian transformation is shown in Eq. 17: In order to solve the optimal solution problem of the above variational constraint, Lagrange multiplier λ is introduced to ensure the strictness of the constraint condition and penalty factor α to ensure the accuracy of signal reconstruction in high noise environment, and transforms the constraint problem into an unconstrained variational problem.

$$\begin{aligned} L_{\mathrm{VMD}}(\{u_\kappa\},\{\omega_\kappa\},\lambda) = {} & \alpha \sum_{\kappa}^{\tilde{K}} \left\| \partial_t \left[\vartheta_\kappa(\ell) \right] e^{-j\omega_\kappa \ell} \right\|_2^2 \\ & + \left\| f_{\mathrm{VMD}}(\ell) - \sum_{\kappa}^{\tilde{K}} u_\kappa(\ell) \right\|_2^2 \\ & + \left\langle \lambda(\ell), f_{\mathrm{VMD}}(\ell) - \sum_{\kappa}^{\tilde{K}} u_\kappa(\ell) \right\rangle \end{aligned} \tag{17}$$

(3) Alternate update: initialize $\{u_\kappa^1\}$, $\{\omega_\kappa^1\}$, λ^1, alternate direction multiplier method solution, and iteratively update u_κ^1, ω_κ^1, λ^1.

$$\begin{cases} \hat{u}_\kappa^{\psi+1} \left(\dfrac{\hat{f}_{\mathrm{VMD}}(\omega) - \sum\limits_{\nu \neq \kappa}^{\tilde{K}} \hat{u}_\nu^{\psi}(\omega) + \dfrac{\hat{\lambda}^{\psi}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_\kappa)^2} \right) \\ \hat{\omega}_\kappa^{\psi+1} = \dfrac{\int_0^\infty \omega |\hat{u}_\kappa(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}(\omega)|^2 d\omega} \\ \hat{\lambda}^{\psi+1}(\omega) = \hat{\lambda}^{\psi}(\omega) + \tau \left(\hat{f}_{\mathrm{VMD}}(\omega) - \sum\limits_{\kappa}^{\tilde{K}} \hat{u}_\kappa^{\psi+1}(\omega) \right) \end{cases} \tag{18}$$

where ψ is the number of iterations; τ is the noise tolerance, $\hat{u}_\kappa^{\psi+1}$, $\hat{\omega}_\kappa^{\psi+1}$ represent the mode function and the center frequency at ψ+1 iteration; $\hat{\lambda}^{\psi+1}(\omega)$ is the Lagrange multiplier value of the VMD algorithm at ψ+1 iteration; $\hat{f}_{\mathrm{VMD}}(\omega)$, $\hat{u}_\kappa^{\psi+1}(\omega)$ and $\hat{\lambda}^{\psi}(\omega)$ represent the $f_{\mathrm{VMD}}(\ell)$, $u_\ell^{\psi+1}(\ell)$ and $\lambda^{\psi}(\ell)$ Fourier transform forms respectively.

(4) Submode output as Eq. 19: according to Eq. 18 determine whether the termination conditions, if not, return to step 3), if satisfied, the Fourier inverse transformation of the last update $\{\hat{u}_\kappa(\omega)\}$, get the set $\{u_\kappa(\ell)\}$ within the time domain range, the final output κ submode signal, the submode function is $u_\kappa(\ell)$, and the corresponding center frequency is ω_κ.

$$\sum_{\kappa}^{\tilde{K}} \left\| \hat{u}_\kappa^{\psi+1} - \hat{u}_\kappa^{\psi} \right\|_2^2 \Big/ \left\| \hat{u}_\kappa^{\psi} \right\|_2^2 < \varepsilon \tag{19}$$

where ε is the judgment accuracy $(\varepsilon > 0)$.

4.5 Model construction

The paper exemplifies a 15-min interval to predict the microgrid's source and load power sequences for the next hour. Due to the long time resolution and insufficient accuracy of numerical weather forecasts, only historical meteorological data is utilized as weather feature input during the selection of input features for the prediction model. This data is combined with historical sequences of microgrid source and load power to collectively form the input matrix. Regarding historical features, five historical similar days with significant correlation to the current

weather and the historical power measurement values from the past 5 days are specifically chosen as historical data (Wang, 2020).

The source and load power in microgrids exhibit strong nonlinearity and non-stationarity characteristics, rendering single predictive model methods limited in both fitting performance and prediction accuracy. To enhance power prediction accuracy, this study drew upon the methods outlined in Yue et al. (2023). Initially, VMD was employed to decompose historical source and load power time series under different weather conditions, yielding multiple IMF components of various frequencies. Subsequently, the permutation entropy (PE) values of each IMF were computed, and based on these PE values, low, medium, and high-frequency input matrices were constructed. Considering the concurrent temporal correlation of current microgrid source and load power values with past and future time information, three homogeneous recursive neural network models—BP, LSTM, and BiLSTM—were selected for their robust handling of time-series data. These models were employed as base learners, utilizing a bootstrapping method to acquire diverse training set samples, which were then used to train the base learners. This approach enabled the prediction of different frequency components, which were subsequently combined to obtain microgrid source and load power forecasts. High-frequency data changes typically exhibit strong sequential dependencies and long-term trends. LSTM models excel at capturing long-term dependencies within sequential memory and adaptively adjusting the complexity and variability of sequence patterns, making them suitable for predicting high-frequency data trends. Medium-frequency data is often influenced by preceding and succeeding time-step data, exhibiting certain contextual dependencies. BiLSTM models, equipped with both forward and backward memory units, can simultaneously process forward and reverse sequence information, thus better capturing contextual relationships within medium-frequency data and enhancing prediction accuracy. Low-frequency component variations are relatively slow and stable. The training process of BP neural networks is relatively straightforward, capable of providing forecasts of future trends by learning the input-output mapping relationships of historical data.

In consideration of the aforementioned, this paper contemplates the correlation between weather characteristics and multivariable factors, proposing a joint prediction model for microgrid source and load power based on MV-UIC-FA. The schematic diagram of the prediction model is illustrated in Figure 5. The specific prediction steps are as follows:

(1) Acquiring meteorological characteristics for the forecasted day involves retrieving weather information strongly correlated with source and load power from historical daily datasets. Similar days are selected to construct datasets under distinct weather types.
(2) Data completion: To address gaps in the source and load datasets, missing values are replenished with the average of six data points before and after the sampling point.
(3) Normalization: Employing the Z-score algorithm for normalization ensures a balanced distribution of data.
(4) Timestamp alignment: Concerning the alignment of timestamps in the dataset, this study will utilize spline interpolation for the alignment operation.
(5) Feature selection: Utilizing MV-UIC, an analysis is conducted on the correlation between weather characteristics and microgrid source/load sequences to filter out weather features closely associated with the prediction task.
(6) Feature dimensionality reduction: Employing the FA method, the selected weather feature sequences are subjected to dimensionality reduction while preserving the fundamental information of the original features.
(7) Offline model training: Constructing input matrices for the prediction model involves integrating the processed features with source and load sequences. Following the training methodology of the combined prediction model, the processed dataset is partitioned into training, validation, and testing sets in a ratio of 7:2:1. Subsequently, offline training is conducted to derive the prediction model.

5 Example analysis

To substantiate the rationality of the jointly proposed ultra-short-term forecasting methodology for microgrid source and load power, it is imperative to concurrently acquire the original data pertaining to weather characteristics, distributed power sources, and load power within the microgrid's geographical domain. The source, load, and weather feature data for the DTU 7K 47-bus system (Baviskar et al., 2021), available on the official website, are comprehensive for the period spanning from 1 January 2015, to 31 August 2015. Accordingly, the simulation testing in this study is conducted utilizing the data from this specific timeframe.

The DTU 7K 47-bus system is an open-source multi-voltage level distribution grid model developed by the Technical University of Denmark (DTU). Named the DTU 7k-Bus Active Distribution Grid Model, it spans three voltage levels and is geographically modeled for network topology. Key features include multi-voltage levels enabling analysis of challenges and opportunities in renewable energy-dependent grids, geographical data-based network topology modeling for real-world grid operation simulation, simulation data derived from weather and measured data, and open accessibility for research and educational purposes. Data primarily sourced from DTU's official data-sharing platform, DTU Orbit, allows researchers access for power system analysis, renewable energy integration studies, grid planning, and operational simulations.

The DTU 7K 47-bus system, depicted in Supplementary Appendix Figure S2, is interconnected with the external grid via the B-0 transformer at the B-9 bus; the system encompasses three wind farms, composed of fourth-generation controllable wind turbines with installed capacities of 12, 15, and 15 MW, respectively. Due to the location of the DTU 7K 47-bus testing system within the Danish territory, meteorological data is sourced from the Danish Meteorological Institute. The selected weather features include reflectance, snow reflectance, high cloud cover, low cloud cover, 2-m relative humidity, snow density, 2-m specific humidity, 10-m wind speed, 30-m wind speed, 50-m wind speed, 70-m wind speed, 100-m wind speed, atmospheric pressure, 2-m temperature, total cloud cover, visibility, 10-m wind direction, and mid-level cloud cover, totaling 18 variables. The sampling, normalization, and offline training data sample

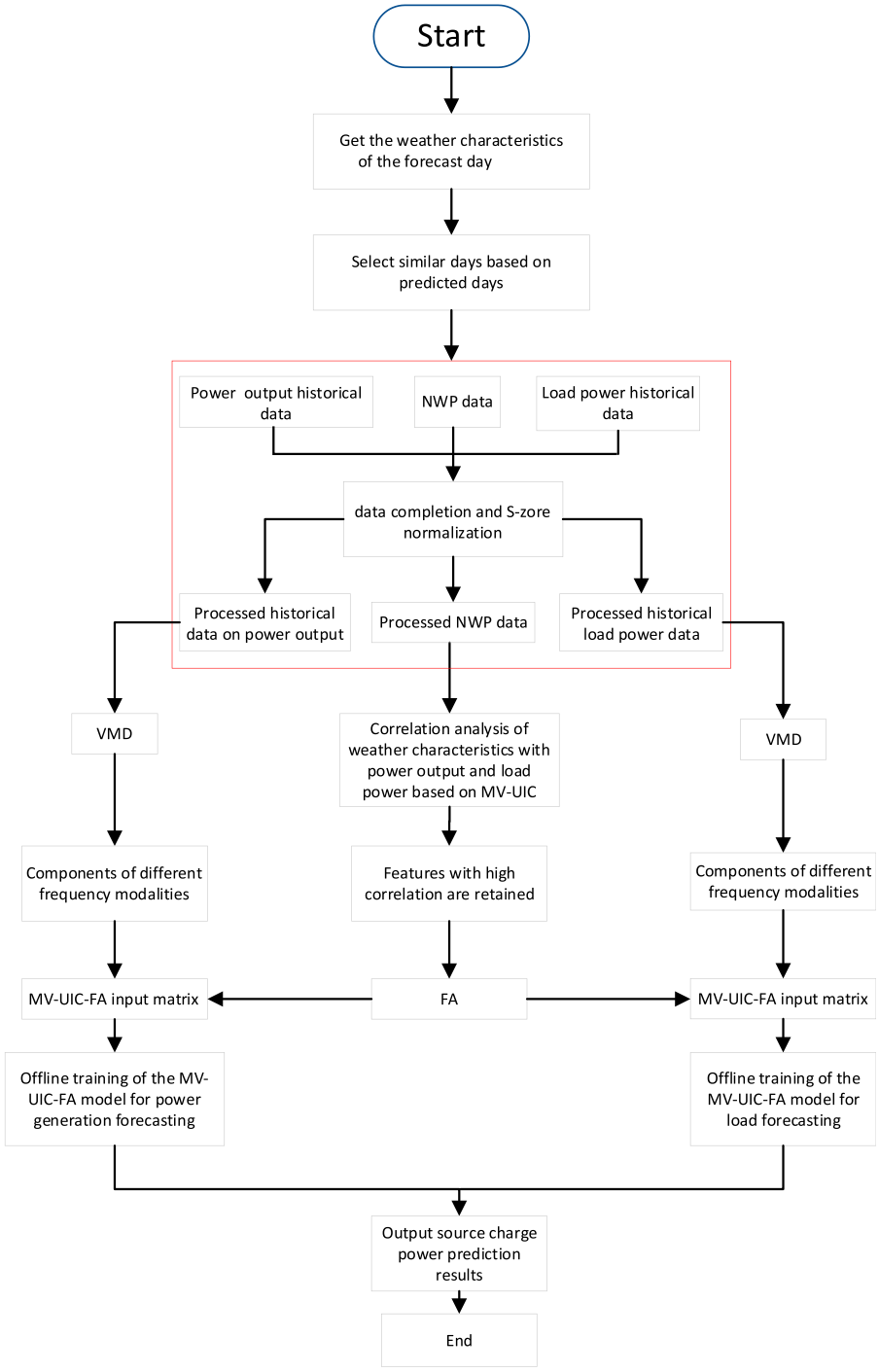


FIGURE 5
Joint prediction model of microgrid source and load power based on MV-UIC-FA.

quantities for source and load data remain consistent. The sampling frequency is 15 min, resulting in a total of 23,328 samples. Z-score algorithm is employed for normalization. Subsequently, 16,330 samples are randomly chosen for training, 4,665 for validation, and 2,333 for testing. The input time series length is set at 96.

The models in this study are trained using the Python software. The performance of the proposed methodology is assessed through the utilization of root mean squared error (RMSE) and mean absolute error (MAE) they are as Eqs 20, 21:

$$v_{\mathrm{RMSE}} = \left(\sum_{i=1}^{N_{\mathrm{T}}} (\hat{v} - v)^2 \Big/ N_{\mathrm{T}} \right)^{\frac{1}{2}} \tag{20}$$

$$v_{\mathrm{MAE}} = \left(\sum_{i=1}^{N_{\mathrm{T}}} |\hat{v} - v| \right) \Big/ N_{\mathrm{T}} \tag{21}$$

TABLE 1 Weather characteristics and MV-UIC of source and load power for four forecast days.

Meteorological characteristics	April 30th (spring)	June 30th (summer)	August 31st (autumn)	February 28th (winter)
Albedo	0.1595	0.1279	0.1932	0.2907
Snow albedo	0.1755	0.0	0.2199	0.3554
High cloud cover	0.1295	0.1719	0.1230	0.1304
Low cloud cover	0.1207	0.1321	0.1234	0.1452
2 m relative humidity	0.1188	0.1449	0.1303	0.1422
Snow density	0.1435	0.0	0.2914	0.3240
2 m specific humidity	0.1965	0.1590	0.1513	0.2382
10 m wind speed	0.1377	0.1586	0.1433	0.2037
30 m wind speed	0.1421	0.1537	0.1405	0.1999
50 m wind speed	0.1314	0.1534	0.1419	0.1966
70 m wind speed	0.1396	0.1608	0.1434	0.1929
100 m wind speed	0.1396	0.1596	0.1462	0.1978
Surface pressure	0.2579	0.2086	0.2985	0.2287
2 m temperature	0.1422	0.1463	0.1333	0.2021
Total cloud cover	0.1335	0.1539	0.1388	0.1564
Visibility	0.1298	0.1399	0.1205	0.1345
10 m wind direction	0.1601	0.1571	0.1518	0.2042
Medium cloud cover	0.1551	0.1541	0.1416	0.1345

In the equation: v is the true value of the source load; $\hat{v}$ is the predicted value of the source load; N_T is the number of elements in the test set.

5.1 Simulation case 1: testing of weather feature extraction and dimensionality reduction methods

5.1.1 Feature extraction and factor analysis based on MV-UIC for dimensionality reduction

To substantiate the rationality of the proposed feature extraction algorithm across diverse seasons, this paper opts for the dates of April 30th (spring), June 30th (summer), August 31st (autumn), and February 28th (winter) as prediction days, aligning with the climatic nuances of Denmark. Historical days with correlation coefficients exceeding 0.8 concerning the prediction days are designated as analogous days. The input feature sequences encompass five historical source-load data points with a kin weather conditions, the source-load data from the past 5 days, and eighteen weather attributes. Four forecasted days' weather features, along with the MV-UIC of source and load power, are extracted as delineated in Table 1.

From Table 1, it can be observed that the weather characteristics vary across different seasons, exhibiting disparate MV-UICs concerning microgrid source and load power. These weather features manifest distinct correlations with source and load power. Given Denmark's temperate maritime climate, precipitation (snow) and strong winds are predominantly observed during the autumn and winter seasons, occasionally culminating in extreme weather phenomena such as blizzards. Consequently, during the winter season, the information coefficients between snow density, wind speed, and source and load power stand at 0.324 and 0.2037, respectively, indicating considerable magnitudes. These findings align with Denmark's actual climatic conditions and geographical location, thereby corroborating the rationality and efficacy of the proposed MV-UIC feature extraction method.

5.1.2 Using FA to reduce the dimension of input features

The FA method is employed herein to capture the common factors among input features by constructing a factor score matrix, facilitating dimensionality reduction for the 18 input features across four forecast days. The total variance explanation table for the spring is presented in Table 2, while the factor score matrix is shown in Table 3. The variance explanation tables and factor score matrices for the other three forecast days can be found in Supplementary Appendix SA. From the variance analysis in Table 2, it is evident that six common factors are extracted from the input features in this study. As indicated by the factor score matrix in Table 3, these factors are the wind speed factor, 2 m information factor (including 2 m relative humidity, 2 m specific humidity, and 2 m temperature), albedo factor, cloud cover factor, surface pressure factor, and wind direction factor. The cumulative variance contribution rate of these common factors is 97.966% (>85%), suggesting that these six

TABLE 2 Total variance interpretation table.

Components	Initial eigenvalues			Extract sum of load			Sum of rotating load		
	Total	Percent variance	Cumulative %	Total	Percent variance	Cumulative %	Total	Percent variance	Cumulative %
1	3.421	38.014	38.014	3.421	38.014	38.014	3.040	33.775	33.775
2	2.013	22.367	60.381	2.013	22.367	60.381	1.787	19.853	53.628
3	1.264	14.040	74.421	1.264	14.040	74.421	1.033	11.478	65.106
4	0.888	9.869	84.290	0.888	9.869	84.290	1.011	11.234	76.339
5	0.661	7.344	91.633	0.661	7.344	91.633	0.981	10.898	87.237
6	0.570	6.332	97.966	0.570	6.332	97.966	0.966	10.729	97.966
7	0.174	1.932	99.898						
8	0.005	0.056	99.955						
9	0.004	0.045	100.000						

TABLE 3 Factor score matrix.

Meteorological characteristics	Factor score					
	Components 1	Components 2	Components 3	Components 4	Components 5	Components 6
Albedo	−0.028	0.149	0.977	−0.070	0.014	0.103
2 m specific humidity	−0.010	0.948	−0.029	0.079	−0.109	0.075
30 m wind speed	0.980	−0.058	−0.016	0.058	0.105	0.131
50 m wind speed	0.980	−0.060	−0.022	0.066	0.104	0.131
70 m wind speed	0.979	−0.060	−0.015	0.065	0.107	0.136
Surface pressure	0.214	−0.142	0.014	−0.110	0.960	−0.032
2 m temperature	−0.131	0.909	0.241	−0.074	−0.056	0.023
10 m wind direction	0.285	0.092	0.114	0.062	−0.034	0.944
Medium cloud cover	0.117	0.011	−0.070	0.983	−0.103	0.056

common factors encompass the majority of the effective information within the sequences of 18 key input features. This is of significant importance for interpreting the variations in the original data.

5.1.3 Test of feature extraction and factor-based dimensionality reduction method based on MV-UIC

To mitigate the non-stationarity of microgrid sources and loads, an initial step involves employing VMD to decompose the time series of source and load data for similar and forecast days. The VMD decomposition results are presented in Supplementary Appendix SB. This process yields various modal sub-sequences of source and load power. Subsequently, the PE values are computed for each sub-sequence. Sequences with PE values exceeding 0.55 are designated as high-frequency sequences, those with PE values ranging from 0.25 to 0.55 are categorized as mid-frequency sequences, and sequences with PE values below 0.25 are identified as low-frequency sequences. The decomposition results of the spring forecast day source and load power are illustrated in Supplementary Appendix Figure SFA1, revealing the absence of low-frequency components in the wind power output sequence.

In order to validate the rationality of the proposed feature extraction algorithm in different seasons, the study utilized the all feature (AF) of the forecast day and input features extracted based on MV-UIC-FA separately as inputs for the prediction model. Three types of base learners—BP, LSTM, and BiLSTM—were employed to individually predict the low, mid, and high-frequency components of source and load power. The corresponding evaluation metrics for the prediction results are presented in Table 4, with a prediction step of 4 for all control groups. Upon examination of Table 4, it is observed that compared to using all weather features as input for the prediction model, employing MV-UIC to extract and dimensionally reduce weather features before prediction resulted in a reduction of 13.46% and 17.85% in RMSE and MAE for electric power prediction, respectively. For load power prediction, the RMSE and MAE

TABLE 4 Evaluation indexes of source and load power prediction models after dimensionality reduction of AF and MV-UIC-FA.

	Input features	April 30 (spring)		June 30 (summer)		August 31 (fall)		February 28 (winter)	
		RMSE/MW	MAE/MW	RMSE/MW	MAE/MW	RMSE/MW	MAE/MW	RMSE/MW	MAE/MW
DG	AF	2.0861	1.7307	1.1229	0.8462	4.5418	3.6165	3.3936	2.9253
	MV-UIC-FA	1.7910	1.4537	0.9885	0.7241	4.067	3.1751	2.7818	2.0841
Load	AF	8.9578	7.6002	14.1083	10.1921	9.0828	7.5761	11.7821	9.2057
	MV-UIC-FA	6.9727	5.7600	11.8506	8.0875	7.5069	6.7075	10.0187	8.0180

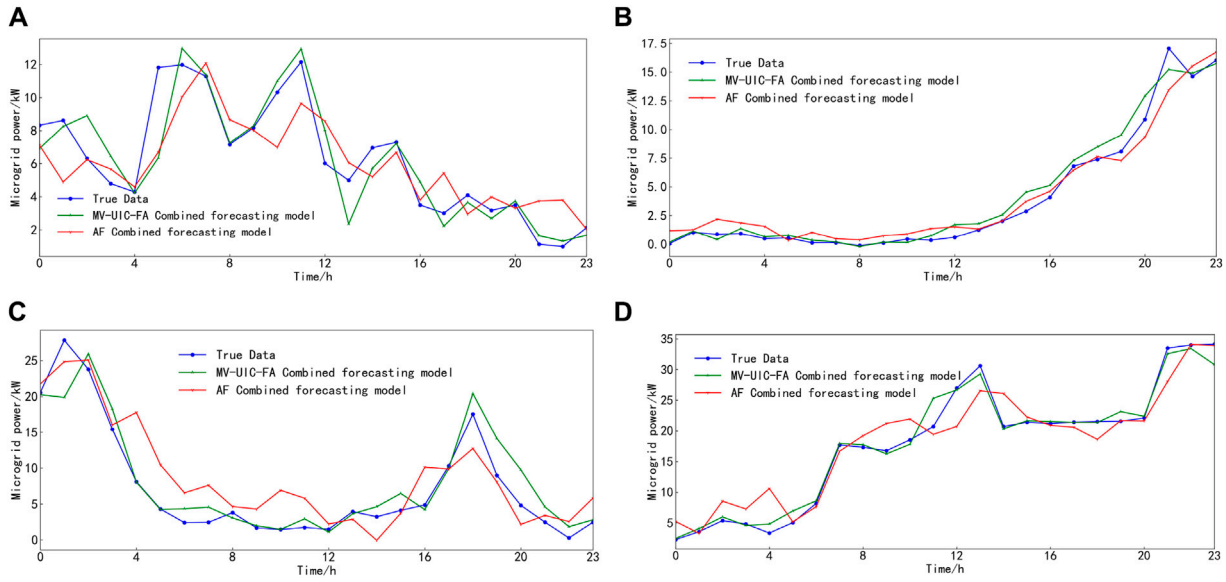


FIGURE 6
Microgrid power prediction results for four prediction days. **(A)** Spring. **(B)** Summer. **(C)** Fall. **(D)** Winter.

were reduced by 17.62% and 17.31%, respectively. This indicates a significant improvement in prediction accuracy, validating the effectiveness of the proposed approach across different seasons.

This simulation yielded the corresponding predictions as shown in Figures 6, 7.

5.2 Simulation case 2: comparative testing of accuracy of different prediction models

To verify the accuracy of the joint prediction model for microgrid source and load power based on MV-UIC-FA proposed in this paper, predictions of source and load power for four forecasting days were initially conducted using the MV-UIC-FA prediction model. Subsequently, comparisons were made with the results of three single prediction models, namely BP (MV-UIC-BP, MV-B), LSTM (MV-UIC-LSTM, MV-L), and BiLSTM (MV-UIC-BiLSTM, MV-Bi). These models utilized dimensionally reduced input features. Evaluation metrics such as RMSE and MAE for the corresponding prediction results were obtained and are presented in Table 5 (with a prediction step of 4 for all control groups). Additionally, comparisons between the predicted results for source and load power for the four forecasting days and the actual data are illustrated in Figures 8, 9.

From Figures 8, 9, it can be observed that among the predicted power for the four forecasting days, the results of the three individual forecasting models are similar, while the combined forecasting model fully exploits the advantages of each individual forecasting model, yielding superior forecasting results. As seen from Table 5, the proposed models in this study outperform various baseline models in terms of their RMSE and MAE. Specifically, compared to the forecasting results of the BP, LSTM, and BiLSTM models, the employment of the proposed model in this study reduces the RMSE and MAE of the power supply by 25.57%, 25.71%, 23.48%, and 33.4%, 31.41%, 26.92%, respectively, and reduces the RMSE and MAE of the load power by 25.69%, 18.69%, 18.33%, and 26.79%, 17.71%, 19.07%, respectively.

5.3 Simulation case 3: comparative testing of accuracy of different prediction models

In this section, the accuracy and robustness of the proposed joint prediction model for microgrid source and load power based on

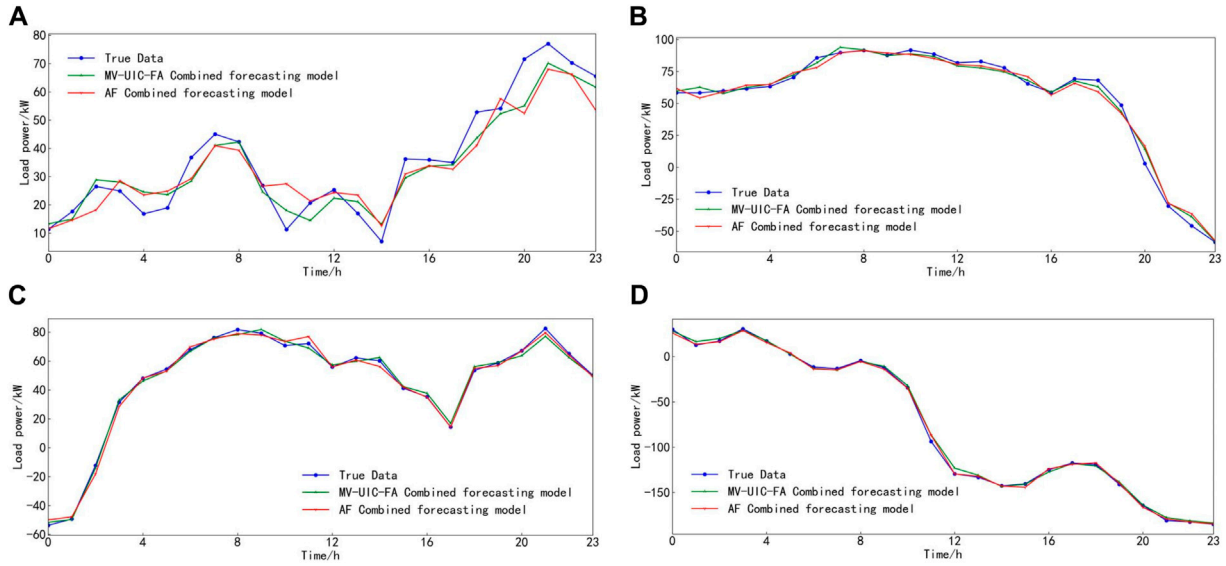


FIGURE 7
Microgrid load prediction results for four prediction days. **(A)** Spring. **(B)** Summer. **(C)** Fall. **(D)** Winter.

TABLE 5 Evaluation indexes of RMSE and MAE for different prediction models.

	Prediction model	Spring		Summer		Fall		Winter	
		RMSE/MW	MAE/MW	RMSE/MW	MAE/MW	RMSE/MW	MAE/MW	RMSE/MW	MAE/MW
DG	MV-B	2.2125	1.9671	0.9573	0.7082	4.2045	3.5404	2.7744	2.2158
	MV- L	2.0732	1.9361	1.1927	0.8617	4.2443	3.3818	2.3355	1.6197
	MV- Bi	2.1235	1.9527	1.0177	0.8208	3.0152	1.876	3.4657	2.5119
	The MV-UIC-VMD combined prediction model	1.7910	1.4537	0.8585	0.6241	2.5884	1.7535	1.8183	1.2153
Load	MV-B	10.8240	8.1104	9.9246	9.7192	21.6300	15.2434	25.2227	22.7800
	MV- L	8.4671	6.3477	9.7702	9.0331	18.9970	14.1968	25.3980	21.1923
	MV- Bi	8.8633	6.7884	7.6545	6.2334	19.0853	13.5598	23.7700	18.2401
	The MV-UIC-VMD combined prediction model	6.9727	5.7600	8.5803	7.9440	14.4447	10.3772	20.0735	16.4130

MV-UIC-FA are validated using the IEEE 118-bus standard distribution network test system (Youssef et al., 2020) as shown in Figure 10. Actual power and load data from the 2014 Global Energy Forecasting Competition (GEFC) (Hong et al., 2016) are utilized as the training and testing datasets. Controllable wind turbines of the fourth generation, each with a rated capacity of 1 MW, are connected to buses 14, 25, 46, 49, 66, and 69 of the IEEE 118- bus test system. Load data and environmental weather characteristic data from the 2014 GEFC are extracted, covering the period from March 1st to 12th, 2005, with a sampling frequency of 1 h, totaling 288 h of data. This dataset is used for experimental simulations. Meteorological data utilized in the simulations are sourced from the publicly available local weather information on the website of the National Renewable Energy Laboratory (NREL) (NREL, 2024) in the United States. The dataset includes load and power sequences, temperature, weather type 1, humidity, visibility, weather type 2, perceived temperature, pressure, wind speed, cloud cover, wind resistance, precipitation intensity, dew point, and precipitation probability. Normalization is applied to both power and load data to meet simulation requirements. Missing values in the dataset are filled, and normalization is performed. The training and prediction datasets are divided in an 8:2 ratio. March 12th is selected as the prediction day.

Based on the findings presented in Figure 11 and Table 6, it is evident that the predictive performance of the method proposed in this paper for forecasting source and load power within the IEEE 118-bus standard distribution network surpasses that of alternative individual methods, demonstrating superior predictive accuracy and commendable generalization capabilities.

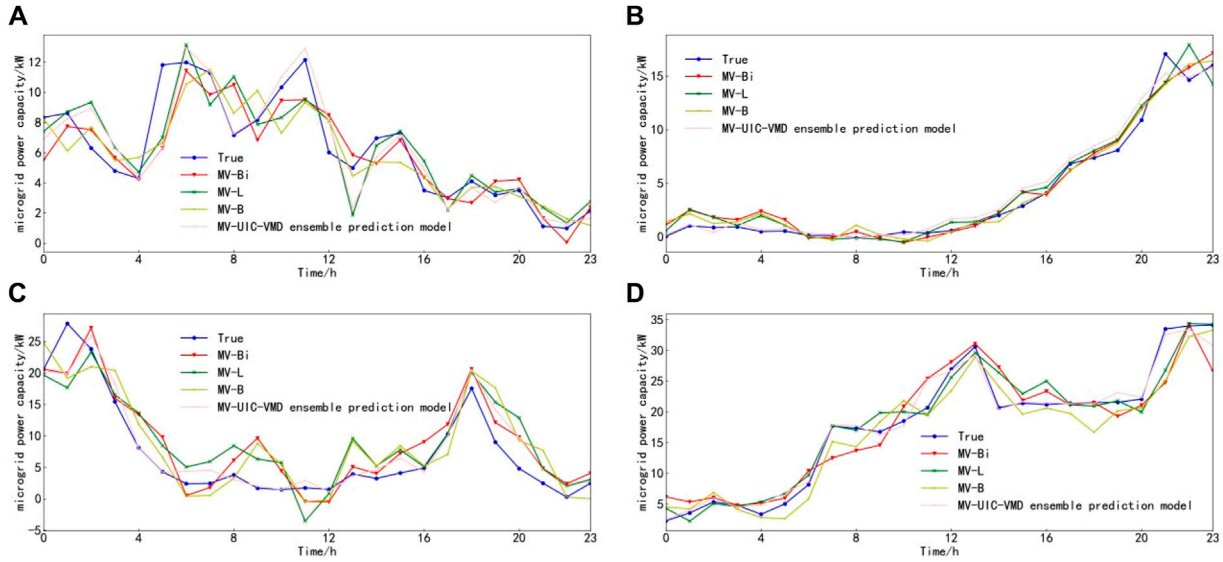


FIGURE 8
Microgrid power prediction results for four prediction days. **(A)** Spring. **(B)** Summer. **(C)** Fall. **(D)** Winter.

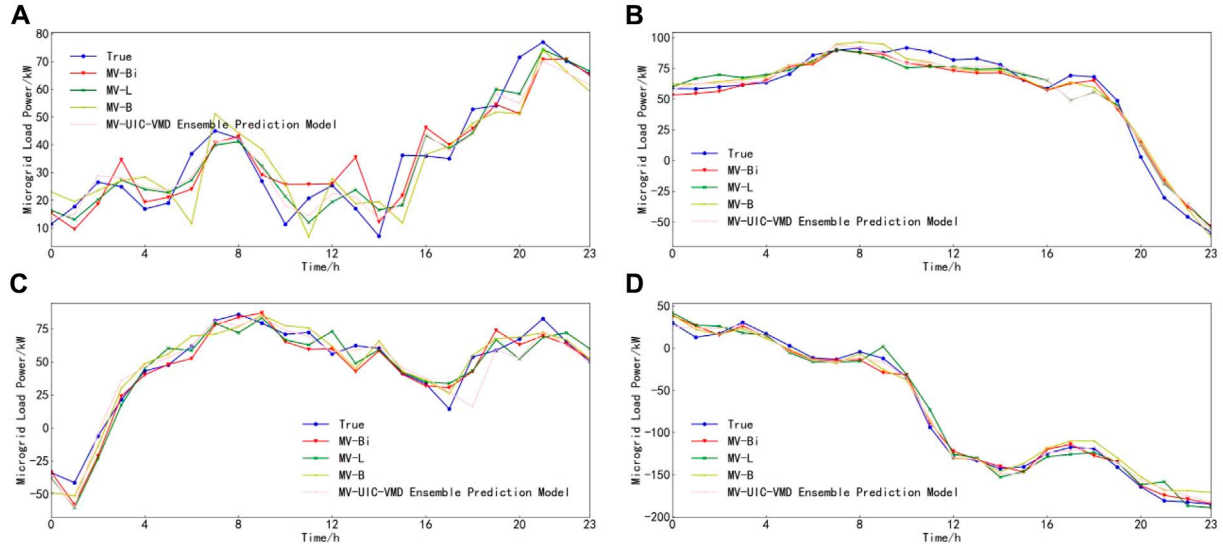


FIGURE 9
Microgrid load prediction results for four prediction days. **(A)** Spring. **(B)** Summer. **(C)** Fall. **(D)** Winter.

5.4 Simulation case 4: with the prediction results that consider only the distributed power supply and load alone in correlation with the weather feature

For comparative analysis, the influence of weather features only on DG output and load power is set separately using the proposed algorithm to predict the load in the above two scenarios. The prediction results are shown in Figures 12, 13 and Table 7.

Based on the simulations of the two scenarios mentioned above, it can be inferred that considering the correlation between DG output, load power, and weather characteristics can further improve the accuracy of load forecasting. Additionally, the proposed algorithm in this paper has a prediction time of 33.18 s, demonstrating good timeliness and meeting the requirements of ultra-short-term load forecasting.

6 Conclusion

In response to the coexistence of distributed power sources and loads in microgrids, wherein weather characteristics

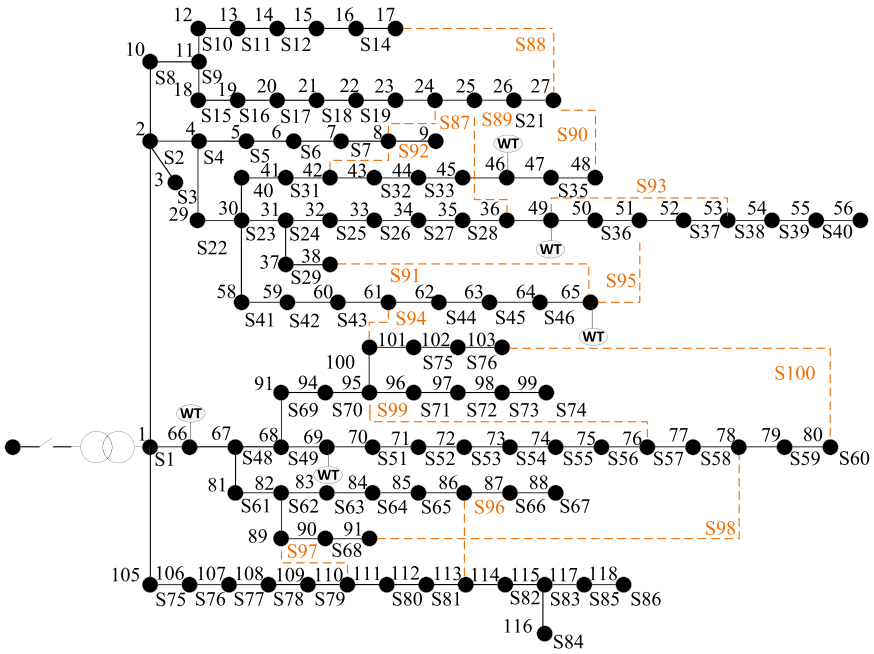


FIGURE 10
IEEE 118-bus standard power distribution system.

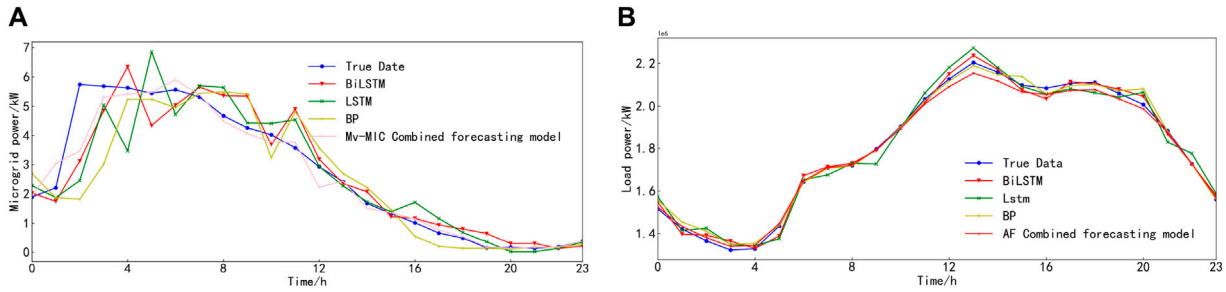


FIGURE 11
Results diagram of source and load power prediction of IEEE 118-bus standard distribution system. **(A)** DG. **(B)** Loads.

TABLE 6 Evaluation indexes of source and load power prediction models after dimensionality reduction of IEEE 118-bus standard distribution system.

	Prediction model	RMSE/MW	MAE/MW
DG	MV-B	1.108	0.6748
	MV-L	0.9564	0.5899
	MV-Bi	0.7693	0.5306
	The MV-UIC-VMD combined prediction model	0.5361	0.2627
Load	MV-B	0.2492	0.1709
	MV-L	0.4066	0.3486
	MV-Bi	0.2497	0.2084
	The MV-UIC-VMD combined prediction model	0.1143	0.1143

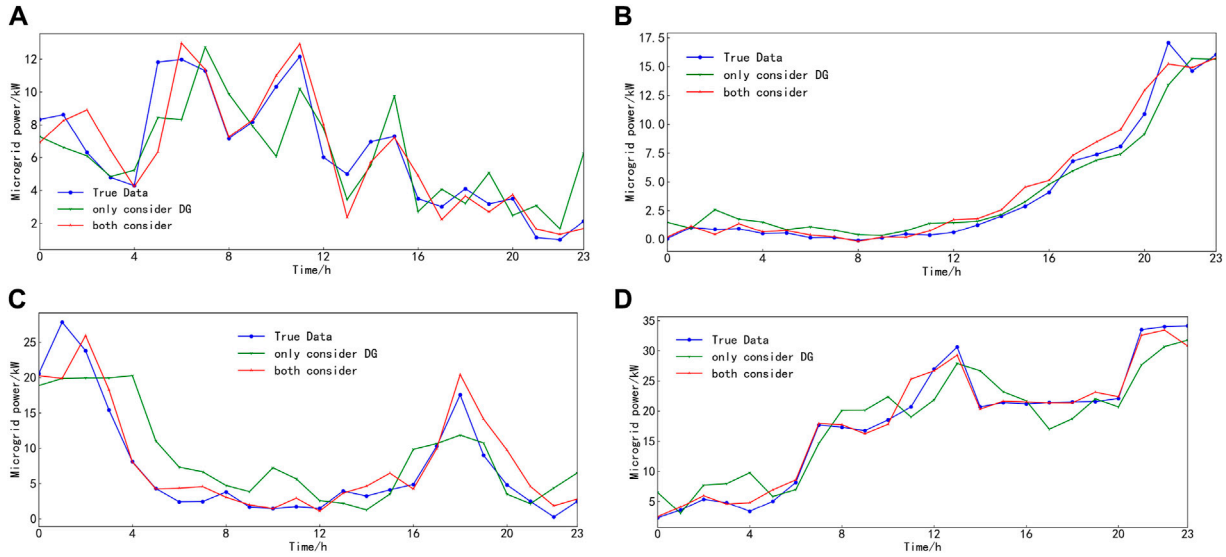


FIGURE 12
Only the predicted results of the DG output were considered. **(A)** Spring. **(B)** Summer. **(C)** Fall. **(D)** Winter.

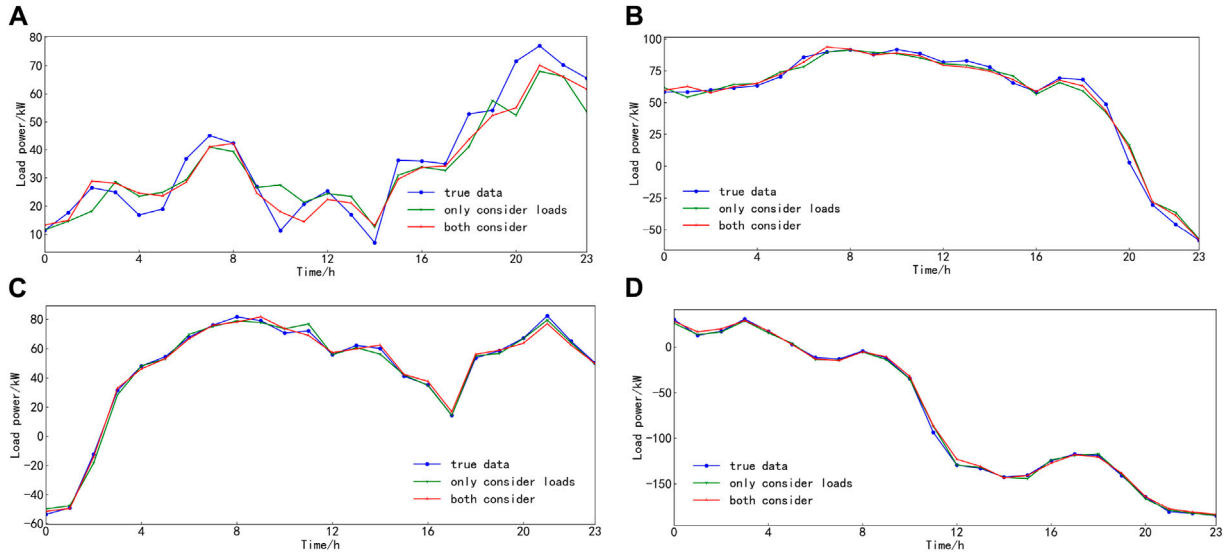


FIGURE 13
Only the predicted results of the loads were considered. **(A)** Spring. **(B)** Summer. **(C)** Fall. **(D)** Winter.

TABLE 7 Evaluation indexes of source and load power prediction models after dimensionality reduction.

Prediction model	Spring		Summer		Autumn		Winter	
	RMSE/MW	MAE/MW	RMSE/MW	MAE/MW	RMSE/MW	MAE/MW	RMSE/MW	MAE/MW
Only consider loads	6.30469	5.00133	5.01234	3.53141	2.44036	2.08708	2.28101	1.84892
Only consider DG	2.0457	1.5801	1.0663	0.8019	3.7107	2.9303	3.1595	2.3205
Both consider	1.5805	1.039	0.8585	0.6241	2.5884	1.7535	1.3818	0.8841

concurrently influence their power, a joint short-term power prediction model for microgrid sources and loads, considering weather features and multivariable correlations, is proposed to attain a rational match between microgrid sources and loads. Illustrated by an analysis of the DTU 7K 47-bus system within Denmark, an assessment of the accuracy, applicability, and efficacy of the proposed prediction approach is conducted. The principal findings are as follows:

(1) MV-UIC can effectively depict the simultaneous impact of the same weather characteristics on the power of sources and loads within microgrids, thereby revealing the correlation between weather features and microgrid power.
(2) By employing MV-UIC in conjunction with factor analysis to reduce the dimensionality of input features for source and load prediction, the power forecasting accuracy surpasses that achieved when considering all weather features as input. Compared to single prediction models, utilizing the prediction model based on MV-UIC-FA for source and load power also effectively reduces prediction errors.

Upon deriving the predicted power for microgrid sources and loads through the methodology advanced in this paper, the subsequent phase will involve the modeling of the matching degree between microgrid sources and loads, coupled with the optimization scheduling research of microgrid clusters.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://orbit.dtu.dk/en/datasets/dtu-7k-bus-active-distribution-network.

Author contributions

ZH: Conceptualization, Methodology, Visualization, Writing–original draft. NS: Supervision, Validation, Writing–review and editing. HS: Project administration, Writing–review and editing. YL: Funding acquisition, Resources, 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

Author ZH was employed by State Grid Shandong Electric Power Company. Authors NS, HS, and YL were employed by State Grid Yantai 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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2024.1409957/full#supplementary-material

References

Baviskar, A., Hansen, A. D., Das, K., and Douglass, P. J. (2021). "Open-source active distribution grid model with a large share of res-features, and studies," in 2021 9th IEEE International Conference on Power Systems (ICPS), Kharagpur, India, 16-18 December 2021 (IEEE), 1–6. doi:10.1109/ICPS52420.2021.9670223

Chang, G. W., and Lu, H. J. (2018). Integrating gray data preprocessor and deep belief network for day-ahead PV power output forecast. *IEEE Trans. Sustain. Energy* 11 (1), 185–194. doi:10.1109/tste.2018.2888548

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., and Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *Int. J. Forecast.* 32 (3), 896–913. doi:10.1016/j.ijforecast.2016.02.001

Hossain, M. S., and Mahmood, H. (2020). Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *Ieee Access* 8, 172524–172533. doi:10.1109/access.2020.3024901

Jiang, F., Lin, Z., Wang, W., Wang, X., Xi, Z., and Guo, Q. (2023). Optimal bagging ensemble ultra short term multi-energy load forecasting considering least average envelope entropy load decomposition. *Proc. CSEE* 43 (08), 3027–3048. doi:10.13334/j.0258-8013.pcsee.223470

Kobayashi, K., and Iwai, M. (2018). Quantitative independent component selection using attractor analysis for noise reduction in magnetocardiogram signals. *IEEE Trans. Magnetics* 54 (11), 1–4. doi:10.1109/tmag.2018.2845903

Lin, Z. (2022). Short-term prediction of building sub-item energy consumption based on the CEEMDAN-BiLSTM method. *Front. Energy Res.* 10, 908544. doi:10.3389/fenrg.2022.908544

Masoumi, A., Jabari, F., Ghassem Zadeh, S., and Mohammadi-Ivatloo, B. (2020). Long-term load forecasting approach using dynamic feed-forward back-propagation artificial neural network. *Optim. Power Syst. Problems Methods, Algorithms MATLAB Codes*, 233–257. doi:10.1007/978-3-030-34050-6_11

Ma, W., Qiao, Y., Lu, Z., Li, J., Sun, S., and Zhou, Q. (2023). Short-term wind power prediction based on combination of screening and optimizing sensitive meteorological characteristics. *Power Syst. Technol.* 47 (07), 2897–2908. doi:10.13335/j.1000-3673.pst.2022.1515

Mousavi, A., and Baranuk, R. G. (2022). Uniform partitioning of data grid for association detection. *IEEE Trans. Pattern Analysis Mach. Intell.* 44 (2), 1098–1107. doi:10.1109/tpami.2020.3029487

Ng, D. T. K., Wu, W., Leung, J. K. L., and Chu, S. K. W. (2023). "Artificial Intelligence (AI) literacy questionnaire with confirmatory factor analysis," in 2023 IEEE International Conference on Advanced Learning Technologies (ICALT), Orem, UT, USA, July 10-13, 2023 (IEEE), 233–235. doi:10.1109/ICALT58122.2023.00074

NREL (2024). *National renewable energy laboratory (NREL) home page | NREL*.

Qun, Y. U., Huo, X., He, J., Li, L., Zhang, J., and Feng, Y. (2023). Trend prediction of power blackout accidents in Chinese power grid based on spearman's correlation coefficient and system inertia. *Proc. CSEE* 43 (14), 5372–5381. doi:10.13334/j.0258-8013.pcsee.220035

Ramirez, D., Santamaria, I., Scharf, L. L., and Van Vaerenbergh, S. (2019). Multi-channel factor analysis with common and unique factors. *IEEE Trans. Signal Process.* 68, 113–126. doi:10.1109/tsp.2019.2955829

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data Sets. *SCIENCE* 334 (6062), 1518–1524. doi:10.1126/science.1205438

Safari, N., Chung, C. Y., and Price, G. (2018). Novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular

spectrum analysis. *IEEE Trans. power Syst.* 33 (1), 590–601. doi:10.1109/tpwrs.2017.2694705

Wang, K., Wang, C., Yao, W., Zhang, Z., Liu, C., Dong, X., et al. (2024). Embedding P2P transaction into demand response exchange: a cooperative demand response management framework for IES. *Appl. Energy* 367, 123319. doi:10.1016/j.apenergy.2024.123319

Wang, T., Liang, H., Cao, J., and Zhao, Y. (2023). Probabilistic power flow calculation using principal component analysis-based compressive sensing. *Front. Energy Res.* 10, 1056077. doi:10.3389/fenrg.2022.1056077

Wang, Y. (2020). *Algorithm analysis and improvement of maximum information coefficient*. Xi'an: University of Electronic Science and Technology.

Wu, C., Sun, S., Cui, Y., and Xing, S. (2024). Driving factors analysis and scenario prediction of CO2 emissions in power industries of key provinces along the Yellow River based on LMDI and BP neural network. *Front. Ecol. Evol.* 12, 1362541. doi:10.3389/fevo.2024.1362541

Xu, W., Shen, Z., Fan, X., and Liu, Y. (2023). Short-term wind power prediction based on anomalous data cleaning and optimized lstm network. *Front. Energy Res.* 11, 1268494. doi:10.3389/fenrg.2023.1268494

Youssef, H. H., Mokhlis, H. B. I. N., Talip, M. S. A., Samman, M. A., Muhammad, M. A., and Mansor, N. N. (2020). Distribution network reconfiguration based on artificial network reconfiguration for variable load profile. *Turkish J. Electr. Eng. Comput. Sci.* 28 (5), 3013–3035. doi:10.3906/elk-1912-89

Yu, Y., Jin, Z., Ćetenović, D., Ding, L., Levi, V., and Terzija, V. (2024). A robust distribution network state estimation method based on enhanced clustering Algorithm: accounting for multiple DG output modes and data loss. *Int. J. Electr. Power & Energy Syst.* 157, 109797. doi:10.1016/j.ijepes.2024.109797

Yue, Y. U., Guo, J., and Jin, Z. (2023). Optimal extreme random forest ensemble for active distribution network forecasting-aided state estimation based on maximum average energy concentration VMD state decomposition. *Energies* 16 (15), 5659–5664. doi:10.3390/en16155659

Zhang, Z., Wang, C., Wu, Q., and Dong, X. (2024). Optimal dispatch for cross-regional integrated energy system with renewable energy uncertainties: a unified spatial-temporal cooperative framework. *Energy* 292, 130433. doi:10.1016/j.energy.2024.130433

Zhou, Y., Qian, C., Wang, Y., Wang, W., Zhou, J., and Wang, H. (2020). Feature parameter extraction of load model identification based on clustering algorithm and class noisy data. *Adv. Technol. Electr. Eng. Energy* 39 (12), 12–18. doi:10.12067/ATEEE200801

Zhu, Q., Jiateng, Li, Qiao, Ji, Shi, M., and Wang, C. (2023). Application and outlook of artificial intelligence technology in new energy power forecasting. *Proc. CSEE* 43 (08), 3027–3048. doi:10.13334/j.0258-8013.pcsee.213114

Zhu, R., Guo, W., and Gong, X. (2019). Short-term load forecasting for CCHP systems considering the correlation between heating,gas and electrical loads based on deep learning. *Energies* 12 (17), 3308–3312. doi:10.3390/en12173308

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 12 June 2024
DOI 10.3389/fenrg.2024.1408119

Check for updates

OPEN ACCESS

EDITED BY
Jinran Wu,
Australian Catholic University, Australia

REVIEWED BY
Yanan Song,
Xi'an Jiaotong University, China
Minh Quan Duong,
The University of Danang, Vietnam

*CORRESPONDENCE
Akanksha Jain,
akanksha091091@gmail.com

RECEIVED 27 March 2024
ACCEPTED 14 May 2024
PUBLISHED 12 June 2024

CITATION
Jain A and Gupta SC (2024), Evaluation of electrical load demand forecasting using various machine learning algorithms.
Front. Energy Res. 12:1408119.
doi: 10.3389/fenrg.2024.1408119

COPYRIGHT
© 2024 Jain and Gupta. 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.

Evaluation of electrical load demand forecasting using various machine learning algorithms

Akanksha Jain* and S. C. Gupta

Electrical Engineering, Maulana Azad National Institute of Technology, Bhopal, India

The energy sector heavily relies on a diverse array of machine learning algorithms for power load prediction, which plays a pivotal role in shaping policies for power generation and distribution. The precision of power load prediction depends on numerous factors that reflect nonlinear traits within the data. Notably, machine learning algorithms and artificial neural networks have emerged as indispensable components in contemporary power load forecasting. This study focuses specifically on machine learning algorithms, encompassing support vector machines (SVMs), long short-term memory (LSTM), ensemble classifiers, recurrent neural networks, and deep learning methods. The research meticulously examines short-term power load prediction by leveraging Chandigarh UT electricity utility data spanning the last 5 years. The assessment of prediction accuracy utilizes metrics such as normalized mean square error (NMSE), root mean squared error (RMSE), mean absolute error (MAE), and mutual information (MI). The prediction results demonstrate superior performance in LSTM compared to other algorithms, with the prediction error being the lowest in LSTM and 13.51% higher in SVMs. These findings provide valuable insights into the strengths and limitations of different machine learning algorithms. Validation experiments for the proposed method are conducted using MATLAB R2018 software.

KEYWORDS

forecasting, power load, machine learning, deep learning, load demand

1 Introduction

Load forecasting serves as a crucial intermediary, ensuring a seamless connection between electricity generation and distribution. Its primary objective is to precisely forecast the electricity load for the upcoming year, months, and weeks, encompassing both short- and long-term projections. Effective power load forecasting enables the efficient management of power distribution scarcity. Demand forecasting also plays a pivotal role in driving nations' industrialization and urban development (Lai et al., 2020; Aslam et al., 2021; Fan et al., 2019; Mosavi et al., 2019). Accurate forecasting is crucial for effective planning and promoting economic growth within a nation. The power load forecasting process relies on archived data and statistical models to predict future trends. However, the nonlinear nature of power generation data frequently leads to increased prediction errors, which can compromise decision-making regarding power generation and distribution. Despite the existence of numerous mathematical models for power load forecasting, attaining high accuracy in these forecasts remains a significant challenge (Su et al., 2019; Khan W. et al., 2020). Predicting the power load is a contemporary research focus. The advancement of machine learning (ML) algorithms propels the evolution of

machine learning and its application in energy forecasting. The amalgamation of sensor technology with power distribution results in the accumulation of a significant volume of data. These amassed data present both opportunities and challenges for making informed decisions. In-depth data processing is conducted for evaluation and forecasting, with machine learning algorithms and models playing a pivotal role in prediction. Due to their efficiency and effectiveness, these algorithms and models have garnered considerable importance in predictive modeling for production, consumption, and demand analysis in recent years (Ahmad W. et al., 2020, Almaghrebi et al., 2020). Despite the extensive research conducted on machine learning and the advancements made in memory-based algorithms, innovative approaches have been proposed for predicting electricity demand. Several statistical functions have been used to model and forecast demand, including gray models, linear regression, autoregressive average models, and partial linear models, all of which are widely utilized in this field (Khan P. et al., 2020; Reynolds et al., 2019; O'dwyer et al., 2019). However, while strong predictive outcomes can generally be achieved, statistical methods are constrained by the underlying linear assumption. The gray prediction model operates without relying on statistical assumptions; nevertheless, its predictive accuracy depends on the dispersion level within the input time series (Chapaloglou et al., 2019; Bedi and Toshniwal, 2019; Jiang et al., 2020). Additionally, due to the distinctive strengths and limitations of each model, it is rare for a single forecasting model to maintain superiority in every situation. Another area of this study focuses on the evolution of load forecasting, transitioning from statistical to hybrid forecasting methods that integrate intelligent approaches capable of addressing complex and nonlinear challenges (Satre-Meloy et al., 2020; Wang R. et al., 2020; Ibrahim et al., 2020; Heydari et al., 2020). The implementation of the incremental approach design incorporates machine learning and artificial neural network algorithms (Ullah et al., 2020; Chammas et al., 2019; Santamouris, 2020; Sun et al., 2020). In contemporary data analysis research, both feed-forward neural networks and recurrent neural networks are extensively used across diverse models to achieve precise power load forecasting. The reliability of these predictions depends on the data processing methods used within the decision system. Among the different algorithms used in predictive models for electricity data, one involves the formation of data subsets in time series. Prevalent preprocessing techniques like singular spectrum analysis, ensemble empirical mode decomposition (EMD), and enhanced whole-ensemble empirical mode decomposition with adaptive noise are applied in analyzing electrical data modeling. These methods aid in establishing a procedural framework for extracting essential information from observed load series to forecast future patterns.

The utilization of EMD-based modeling has demonstrated success in managing electricity demand sequences. The research findings suggest that the EMD framework shows promise for accurately forecasting energy demand within specific intervals. Additionally, they developed feature selection techniques to improve the model's performance. They also created a hybrid feature selection technique to extract fundamental knowledge from electricity time series. This finding underscores the importance of utilizing preprocessing techniques to improve predictions.

In addition to data, weather conditions significantly influence the precision of power load forecasting. Integrating various approaches and incorporating energy source guidelines lead to a new algorithm for short-term load prediction, significantly enhancing accuracy (Hu et al., 2020; Prado et al., 2020; El-Hendawi and Wang, 2020). This study investigated an approach to extracting date-associated details from observed load sequences and developed techniques for selecting features to enhance model effectiveness. Moreover, it introduced a hybrid technique to extract essential knowledge from electricity time-series data. It underscores the significance of preprocessing methods in refining predictions. Weather conditions can also impact power load forecasting, introducing complexities due to seasonal effects and reducing the accuracy of specific models. Consequently, many researchers suggest integrating seasonal pattern-effect models into predictive modeling to address this issue (Li et al., 2020; Qiao et al., 2020; Bakay et al., 2021).

The primary objective of this paper is to explore power load forecasting using artificial intelligence methods such as machine learning and artificial neural networks. The research aims to conduct experimental analyses on datasets using various machine learning algorithms to establish a model design for power load forecasting.

The remainder of the paper is structured as follows: Section II concentrates on recent advancements in power load forecasting; Section III outlines the machine learning approaches used for forecasting; Section IV offers an in-depth examination of the experimental methodology; and finally, Section V concludes the paper and offers insights into future directions for further exploration.

2 Related work

In the field of renewable energy forecasting, Lai et al. (2020) conducted a survey and evaluation of machine learning algorithms. Moreover, this work clarified the methodologies utilized in machine learning models for predicting sustainable energy sources, encompassing data preprocessing methods, attribute selection strategies, and performance assessment metrics. Additionally, the study scrutinized renewable energy sources, mean absolute error (MAE) percentages, and coefficients of determination. Aslam et al. (2021) provided a thorough examination of existing deep learning (DL)-based solar modules and wind turbine power forecasting approaches, along with a significant amount of data on electric power forecasting. The study included datasets used in training and validating various predictive models based on deep learning, facilitating the selection of appropriate datasets for new research projects. Fan et al. (2019) developed a novel short-term load prediction algorithm with improved accuracy using the weighted k-nearest neighbor technique. The forecast inaccuracies of this model are juxtaposed against those of the back-propagation neural network model and the autoregressive moving average (ARMA) model. Evaluation through correlation values demonstrates the capability of the proposed forecasting model to offer adaptable advantages, making it suitable for short-term demand forecasting. Mosavi et al. (2019) presented the current state of energy machine learning models, along with a new edition and application taxonomy. A novel methodology is used

to identify and categorize machine learning models based on the method of machine learning simulations, the type of energy, and the application sector. Through the utilization of hybrid machine learning models, the efficiency, resilience, reliability, and generalization performance of machine learning models in energy systems have all significantly improved. Almaghrebi et al. (2020) utilized a dataset obtained from public charging stations over a span of 7 years in Nebraska, United States. The XGBoost regression model outperforms other techniques in forecasting charging requirements, showcasing an RMSE of 6.7 kWh and an R^2 of 52%. Reynolds et al. (2019) presented two approaches to improve district energy management. A heater set-point temperature is implemented to regulate building demand directly. Additionally, it assists in enhancing district heat production through a multi-vector energy hub. These observations underscore the potential benefits of comprehensive energy management, encompassing diverse energy vectors while considering both supply and demand aspects. Khan W. et al. (2020) presented machine learning techniques used to construct a hybrid power forecasting model. Extreme boosting, subcategory boosting, and the random forest (RF) technique are the four machine learning algorithms used. Our hybrid model enhances forecasting by employing feature extraction to preprocess data. While machine learning algorithms are frequently effective in handling high-energy situations, our hybrid version improves forecasts by utilizing feature engineering to preprocess data. Ahmad W. et al. (2020) proposed a groundbreaking deep learning-based technique for forecasting electrical loads. Additionally, a three-step model is developed, incorporating a hybrid feature selector for feature selection, a feature extraction technique to reduce redundancy, and improved support vector machines (SVMs) and extreme learning machines (ELMs) for classification and forecasting. Numerical simulations are graphed, and statistics are presented, suggesting that our upgraded methods are more accurate and perform better than state-of-the-art approaches. Bedi and Toshniwal (2019) proposed acquiring season-based segmentation data and developed a deep learning method for projecting electricity usage while accounting for long-term historical dependency. First, the monthly electricity use data are utilized to conduct cluster analysis. Subsequently, load trends are characterized to enhance the comprehension of the metadata encompassed within each cluster. Jiang et al. (2020) provided forecast intervals that represent the intricacies involved in the design and functioning of power systems with better accuracy. The findings suggest that the suggested model demonstrates encouraging forecasts compared to alternative combined methodologies, which can be advantageous for policymakers and public organizations aiming to maintain the security and stability of the energy infrastructure. Wang R. et al. (2020) proposed a useful enhancement integration-model stacking structure designed to address increasing energy needs. To ensure the comprehensive observation of datasets from diverse spatial and structural perspectives, the stacking model harnesses the strengths of multiple base prediction algorithms, transforming their outcomes into "meta-features." With accuracy gains of 9.5% for case A and 31.6%, 16.2%, and 49.4% for case B, the stacking method outperforms earlier models. Heydari et al. (2020) proposed an innovative and accurate integrated model designed for short-term load and price forecasting. This comprehensive package incorporates the gravitational search algorithm, variational mode decomposition, mixed data modeling, feature selection, and generalized regression neural networks. The proposed model surpasses current benchmark prediction models in terms of precision and stability, as indicated by the findings. Wang et al. (2019) analyzed the systems in depth for forecasting renewable energy based on deep learning methodologies to determine their efficacy, efficiency, and relevance. Additionally, to improve forecasting accuracy, various data preparation strategies and mistake post-correction processes are examined. Several deep learning-based forecasting algorithms are thoroughly investigated and discussed. Sun et al. (2020), in their comprehensive examination, delved deeply into forecasting energy use in buildings. Their meticulous analysis encompasses feature manipulation, potential data-centric models, and projected outcomes, thus encompassing the entirety of the data-driven procedure. In a research project, Ahmad and Chen (2019a) explored short-term energy demand predictions at the district level. They employed two distinct deep learning models. These DL models exhibited higher predictive accuracy at distinct hidden neurons, attributed to the suggested network layout. Hu et al. (2020), utilizing a novel augmented optimization model, constructed and refined it using a differential evolution methodology developed through the bagged echo state network approach. Bagging, a network generalization technique, enhances network generalization while reducing forecasting errors. The suggested model, known for its high precision and reliability, proves to be a valuable method for predicting energy consumption. Walker et al. (2020) explored various machine learning algorithms across a spectrum to estimate the electricity demand in hourly intervals, both at the individual building and aggregated levels. Upon factoring in processing time and error accuracy, the results revealed that random forest and artificial neural network (ANN) models yielded the most accurate predictions at an hourly granularity. Prado et al. (2020) employed methodologies such as the fuzzy inference system model, auto-regressive integrated moving average, support vector regression, adaptive neuro-fuzzy inference system, ANN, ELM, and genetic algorithm. In a sample study, compared to leading artificial intelligence and econometric models, the proposed method attained a 22.3% reduction in the mean squared error and a 33.1% decrease in the mean absolute percentage error. Ahmad et al. (2020a) reported that utility companies require a stable and reliable algorithm to accurately predict energy demand for multiple applications, including electricity dispatching, market involvement, and infrastructure planning. The forecasting results assist in enhancing and automating predictive modeling processes by bridging the gap between machine learning models and conventional forecasting models. El-Hendawi and Wang (2020) introduced the whole wavelet neural network methodology, an ensemble method that incorporates both the overall wavelet packet transform and neural networks. This approach utilizes both components effectively. The proposed methodology has the potential to assist utilities and system operators in accurately predicting electricity usage, a critical aspect for power generation, demand-side management, and voltage stability operations. Zhang et al. (2021) proposed utilizing machine learning approaches for load prediction within the framework of machine learning. The objective is to accomplish tasks through performance measures and

learning from past experiences. They conclude with a list of both well-studied and under-explored sectors that warrant further investigation. The research introduces a neural network model specifically designed to predict short-term loads for a Colombian grid operator spanning a week. The model employs a long short-term memory (LSTM) recurrent neural network and historical load data from a specific region in Colombia. The performance of the model is evaluated using the regression metric MAPE, with the most accurate week displaying an error rate of 1.65% and the least accurate week exhibiting an error rate of 26.22% (Caicedo- Vivas and Alfonso-Morales, 2023). This research introduces a method for reconstructing input features utilizing the maximum information coefficient (MIC). The procedure commences by categorizing load curves through distributed photovoltaic systems (DPVSs) with Gaussian mixture model (GMM) clustering. The presented case study illustrates how this proposed feature reconstruction method significantly enhances the prediction accuracy of deep neural networks (Zheng et al., 2023). The study suggests utilizing long short-term memory Bayesian neural networks for forecasting household loads, especially in scenarios involving EV charging. The findings demonstrate a comparable level of accuracy to point forecasts, coupled with the advantage of providing prediction intervals (Skala et al., 2023).

Pawar and Tarunkumar (2020), within a smart grid featuring a significant renewable energy presence, recommended using an intelligent smart energy management system (ISEMS) to meet energy demands. To achieve accurate energy estimations, the proposed approach compares various prediction models, focusing on both hourly and daily planning. Among these models, the particle swarm optimization (PSO)-based SVM regression model demonstrates superior performance accuracy. Fathi et al. (2020) showed how change impacts the energy efficiency of urban structures using machine learning methods and future climate simulations. Due to the absence of a globally applicable metric for this assessment, determining the most reliable machine learning-based forecast requires an optimal combination of criteria. Somu et al. (2021) suggested that KCNN-exact LSTM holds promise as a deep learning model for forecasting energy demand owing to its capability to recognize spatial and temporal associations within the dataset. To evaluate its dependability, the KCNN-LSTM model was compared against the k-means variant of established electricity usage-pattern forecast models using recognized quality criteria. Ahmad and Chen (2019b), by using genuine pollution data and sustainable consumption records, applied NARM, LMSR, and LS Boost methodologies to forecast the energy demands of large-scale urban utilities, utility firms, and industrial customers. Throughout the summer, fall, winter, and spring periods, the LS Boost model showcased coefficients of variation of 5.019%, 3.159%, 3.292%, and 3.184%, respectively. Ahmad and Chen (2020) conducted a thorough assessment and compared several simulations to select the best forecasting model for obtaining the required result in a number of situations. With coefficients of correlation of 0.972 and 0.971, respectively, the Bayesian regularization backpropagation neural networks and Levenberg Marquardt backpropagation neural networks provide better forecasting accuracy and performance. Ahmad et al. (2020b) used renewable energy and electricity projection models as a key and systematic energy planning tool. The forecast periods are segmented into three separate classifications: short-range, intermediate-range, and long-range. The outcomes of this study will aid practitioners and researchers in recognizing prediction methodologies and selecting relevant methods for achieving their desired goals and forecasting criteria. Choi et al. (2020) developed, in response to the recent power demand patterns, a unique load demand forecasting system constructed using LSTM deep learning techniques. They performed examinations to gauge the inaccuracies of the forecasting module and unexpected deviations in the energy usage patterns within the real-time power demand monitoring system. Su et al. (2019) investigated the ANN, SVM, gradient boosting machines (GBM), and Gaussian process regression (GPR) as examples of data-driven predictive models for natural gas price forecasting. To train the model, quarterly Henry Hub natural gas market pricing data and a pass approach are utilized. These two machine learning algorithms operate differently in predicting natural gas prices, with the ANN demonstrating superior prediction accuracy over the SVM, GBM, and GPR, according to the data. Khan P. et al. (2020) proposed utilizing a variety of data mining approaches, such as preprocessing past demand data and analyzing the properties of the load time series, to examine patterns in energy usage from both renewable and non-renewable energy sources. O'dwyer et al. (2019) investigated recent advancements in the smart energy sector, focusing on methodologies in key application areas and notable implemented examples. They also highlight significant challenges in this sector while outlining future prospects. The aim of this inquiry is to assess the current state of computational intelligence in smart energy management and provide insights into potential strategies to overcome current limitations. Chapaloglou et al. (2019) proposed that smoother diesel generator performance can be achieved by combining it with peak shaving using renewable energy. This approach aims to reduce the demand variability that conventional units must meet. The operation seeks to limit the maximum capacity of diesel engines while simultaneously increasing the supply of renewable energy to the grid. Satre-Meloy et al. (2020) applied a unique dataset containing significant strength and tenant time-use data from United Kingdom homes. They also utilized a groundbreaking clustering approach to capture the entire structure. The discussion focuses on how a customized strategy tailored to the highest demand in residential areas can lead to reductions in demand and mitigation actions. Additionally, it enhances our understanding of the limitations and possibilities for demand flexibility in the household sector. Ibrahim et al. (2020) reported that the increasing interest in machine learning technologies underscores their effectiveness in tackling technological challenges within the smart grid. However, certain hurdles, such as efficient data collection and the examination of intelligent decision-making in complex multi-energy systems, as well as the need for streamlined machine learning-based methods, remain unresolved. Ullah et al. (2020) provided comprehensive insights into the utilization of previous advancements in intelligent transportation systems (ITSs), cybersecurity challenges, the effective use of smart grids for energy efficiency, optimized deployment of unmanned aerial vehicles to enhance 5G and future communication services, and the integration of smart medical systems within the framework of a smart city. Chammas et al. (2019) proposed that LR, SVM, GBM, and RF are four alternative classification algorithms compared to our methodology. A multilayer perceptron (MLP)-based system for calculating the energy consumption of a building based on data from a wireless sensor network (WSN), including luminosity, day of the week, moisture, and temperature, significantly influences the

outcomes observed in the testing set. This yields cutting-edge outcomes with a coefficient of determination R^2 of 64% , RMSE of 59.9%, MAE of 27.3%, and MAPE of 28.04%. Santamouris (2020) reported on energy, peak electricity usage, air pollution, mortality, morbidity, and urban susceptibility. The study also examined recent data on the characteristics and extent of urban overheating, as well as analyses of recent research on the connection between urban heat islands and increasing temperatures. Li et al. (2020) estimated that an SVM and an upgraded dragonfly algorithm are utilized to generate short-term wind electricity forecasts. To enhance the performance of the standard dragonfly approach, an adaptive learning multiplier and a convex optimization strategy are proposed. The suggested model outperforms existing methods, such as MLP networks and Gaussian process models, in terms of forecast precision. Lu et al. (2019) proposed that residential management systems utilize an hour-ahead load management algorithm. A stable pricing methodology derived from artificial neural networks is suggested to address the complexities of future pricing. Calculations involving non-shiftable, shiftable, and guided loads are used to validate the performance of the suggested energy management method (Fathi et al., 2020). Wang H. et al. (2020) carried out a classification study using AI algorithms and current solar power prediction models. Taxonomy is a system for classifying solar energy forecasting methods, optimizers, and frameworks based on similarities and differences. This study can aid scientists and engineers in conceptually analyzing various solar forecast models, allowing them to select the most appropriate model for any given usage scenario. Ghoddusi et al. (2019) proposed that in energy economics publications, SVMs, ANNs, and genetic algorithms (GAs) are among the most commonly utilized methodologies. They explored the successes and limitations of the literature. Gao et al. (2019) provided a prediction method based on the weather conditions of previous days for optimal weather conditions. According to a study of predictive accuracy between new methods and known algorithms, the RMSE accuracy of the predicting approach that is built upon LSTM networks can achieve 4.62%, specifically under ideal weather conditions. Xue et al. (2019) proposed the ability to forecast the best weather conditions; here, a method based on the previous-day climatic data is used. The RMSE accuracy of LSTM infrastructure-predicting approaches can reach 4.62% for favorable climatic circumstances, according to research on the projected accuracy between innovative approaches and known algorithms. Qiao et al. (2020) presented a hybrid approach for carbon dioxide emission forecast that combines the lion swarm optimizer with the genetic algorithm to improve the traditional least squares support vector machine model. When compared with eight previous methods, the novel algorithm demonstrates superior global optimization capabilities, quicker convergence, enhanced accuracy, and moderate computational speed. Bakay et al. (2021) reported that measurements of CO_2, CH_4, N_2O, F-gases, and overall GHG emissions from the energy-generating industry can be predicted using DL, SVM, and ANN approaches. All of the algorithms tested in the study, according to the findings, yielded individually favorable outcomes in predicting GHG emissions. The greatest R^2 value for emissions, according to the expected data, ranges from 0.861 to 0.998, and all conclusions are considered "excellent" regarding the RMSE. Zhou et al. (2019), in their analysis, comprehensively assessed prior driving prediction techniques, highlighting suitable application scenarios for each prediction model. Moreover, it outlines methods to address prediction inaccuracies, aiding designers in selecting suitable driving prediction techniques for varied uses and improving the efficiency of predictive energy management strategies for hybrid and plug-in hybrid electric vehicles. Hao et al. (2019) introduced the DE clustering technique, derived from fundamental morphological processes, to recognize days sharing analogous numerical weather prediction data with the envisaged day within the suggested approach. The progressive generalized regression neural network (GRNN) prediction framework rooted in the DE clustering technique demonstrates superior efficacy in forecasting wind power for the following day compared to the models utilizing DPK clustering-GRNN, AM-GRNN, and K-means clustering-GRNN. Ahmed et al. (2020) discovered that artificial neural network ensembles are the best for generating short-term solar power forecasts, that asynchronous sequential extreme learning machines are the best for adaptive networks, and that the bootstrap procedure is the best for assessing uncertainty When paired with hybrid artificial neural networks and evolutionary algorithms, the findings bring up new possibilities for photovoltaic power forecasting. Antonopoulos et al. (2020) provided a look at how AI is employed in disaster recovery applications. The study categorizes research based on the AI/ML algorithms employed and their applications in energy DR. It culminates by summarizing the strengths and weaknesses of the AI algorithms applied in diverse DR tasks, along with proposing avenues for future research in this burgeoning field (Fathi et al., 2020). Shaw et al. (2019) presented the predictive anti-correlated placement algorithm as a revolutionary algorithm that improves CPU and bandwidth usage. It relies on a comparative analysis of the most commonly utilized prediction models, placed alongside each other for comparison. The practical outcomes illustrate that the suggested approach conserves 18% of energy while reducing service violations by more than 34%, in contrast to several frequently used placement algorithms. Hou et al. (2021) analyzed the impact on energy production, demand, and greenhouse gas emissions. Climate scenario representative concentration pathways (RCPs) are used to project changes in weather elements because of this. Taking into account scenarios RCP2.6, RCP4.5, and RCP8.5, hydro-power production is anticipated to increase by approximately 2.765 MW, 1.892 MW, and 1.219 MW, respectively, in the foreseeable future. Furthermore, the projections suggest a subsequent increase to approximately 3.430 MW, 2.475 MW, and 1.827 MW, respectively. Jørgensen et al. (2020) discovered distinct characteristics in neural networks and support vector machines, which, if modified incorrectly, will cause mistakes. The algorithms can be adjusted to match a variety of situations owing to the many parameters. A growing trend involves utilizing machine learning to digitize wind power estimations.

2.1 Research gap

This project addresses the shortcomings observed in current research, outlined as follows:

Despite the considerable potential offered by ML and DL algorithms, the inherent variability among different techniques remains unexplored. Many investigations focus solely on LSTM, SVM, and EM without comparing their effectiveness against traditional deep learning approaches.

Furthermore, several studies fail to consider methods that showcase the resilience of assessed models, with cross-validation being underutilized. Consequently, the results presented often exhibit significant dependency on the specific data sample used, thereby limiting reproducibility and applicability to future datasets.

3 Methodology

3.1 Data preprocessing

Various machine learning methodologies heavily depend on the caliber and arrangement of the dataset. Implementing efficient preprocessing techniques, which encompass variable selection, data filtration, and transformation into an understandable format for the models, holds paramount significance. Throughout the data collection phase, inaccuracies in communication or gathering frequently result in absent values in the final dataset. Moreover, monitoring programs often capture a myriad of parameters, not all of which contribute to accurately predicting the target variable. To discern the most suitable dataset, a feature extraction procedure is utilized. This involves visually delineating the curves of various variables influencing the target feature through plotting. Additionally, this phase enables the extraction of information not explicitly present as variables but impacting the variability of the target feature, such as the hour of the day, day of the week, or day of the year. Typically, the chosen dataset encompasses weather conditions and electrical load data, with temporal parameters including the hour of the day, day of the week, and day of the year. Consequently, the objective is to forecast the output variable, electrical loads, based on the input variables selected due to their inherent correlation with variations across different time intervals. Specifically, the hour of the day facilitates the extraction of daily patterns, the day of the week reveals weekly patterns, and the day of the year aids in recognizing seasonal patterns.

Improving pattern recognition from the models is accomplished through a data-cleansing phase. Here, the data undergo filtration to identify outliers. Given the substantial instantaneous variability in electricity consumption, the methodology predominantly relies on the interquartile range. This iterative process aims to alleviate errors introduced by anomalous values in the trends. Consequently, values deviating outside the ranges defined by specific criteria are deemed invalid.

$$ub = Q3 + 1.5 * IQR,$$
$$lb = Q1 - 1.5 * IQR,$$

where “ub” and “lb” represent the upper bound and lower bound, respectively; “Q3” and “Q1” denote the third and the first quartiles, respectively; and “IQR” signifies the interquartile range. Values recognized as outliers, in addition to non-existent values within the dataset, are regarded as absent. Data preprocessing serves as the initial stage in machine learning, involving the transformation or encoding of data to prepare them for efficient analysis by the machine. Essentially, this process ensures that the data are in a format that enables the model algorithm to effectively interpret their features.

Data preprocessing holds significant importance for the generalization performance of supervised machine learning algorithms. As the dimensionality of the input space increases, the volume of training data increases exponentially. It is estimated that preprocessing tasks can consume up to 50%–80% of the overall classification process time, underscoring their critical role in model development. Enhancing data quality is also essential for optimizing performance.

The detailed steps of data preprocessing are outlined below.

3.1.1 Data cleaning and validation

Data cleansing involves the identification and rectification or removal of incorrect or noisy data from the dataset. It typically focuses on detecting and replacing incomplete, inaccurate, irrelevant, or other erroneous data and records. Duplicates can frequently occur in datasets, particularly when combining data from various sources, scraping data, or aggregating data from multiple clients. This situation presents an opportunity for the generation of duplicate data.

It is common for certain columns in a dataset to have missing values, which can arise from data validation rules or data collection processes. However, addressing missing values is essential, as they can impact the efficacy of the features of a model. When a significant number of values are missing, straightforward interpolation methods can be used to address these gaps. One of the most prevalent approaches involves using mean, median, or mode values based on the features of the model.

Missing data may result from human error or be generated while working with primary data. Therefore, it becomes necessary to have a data assessment process to learn the datatype of the feature and ensure that all data objects are of the same type. Inconsistent data might lead to erroneous conclusions and forecasts.

3.1.2 Regression (noise handling)

If noise persists within a class even after identifying loud occurrences, there are three strategies for addressing it. First, noise can be disregarded if the model exhibits robustness against overfitting. Second, noise in the dataset can be filtered out, adjusted, refined, or re-labeled. If the attribute-related noise persists, methods such as filtering or refining the erroneous attribute value, excluding it from the dataset, or utilizing imputation techniques can help identify areas requiring cleaning and unveil additional questionable values. This supervised machine learning technique is used for predicting continuous variables by establishing relationships between variables and estimating how each variable influences the others. To assess the predictions made by regression algorithms, it is essential to consider variance and bias metrics.

3.1.3 Data integration

Data integration refers to the amalgamation of data sourced from multiple origins into a unified dataset. This encompasses schema integration, which entails merging metadata from diverse sources and addressing discrepancies in data values stemming from variations in units of measurement, representation, and other factors. Additionally, it is essential to manage redundant data by employing techniques such as correlational analysis to uphold high data quality post-integration.

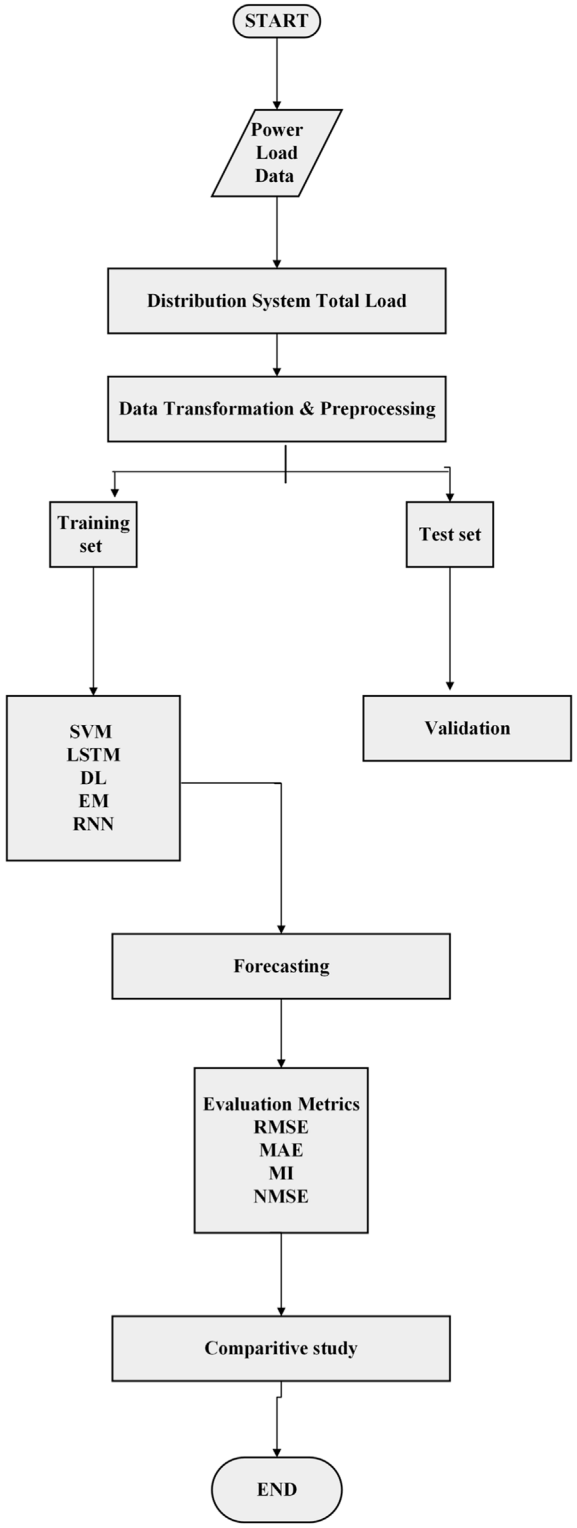


FIGURE 1
Flowchart for power load forecasting with machine learning algorithms.

3.1.4 Data transformation (normalization)

Normalization becomes necessary when attributes are measured on different scales. In cases where multiple features exhibit distinct scales, normalization is essential to standardize them or risk yielding suboptimal outcomes. This process encompasses techniques such as min–max normalization, z-score normalization, and decimal scaling.

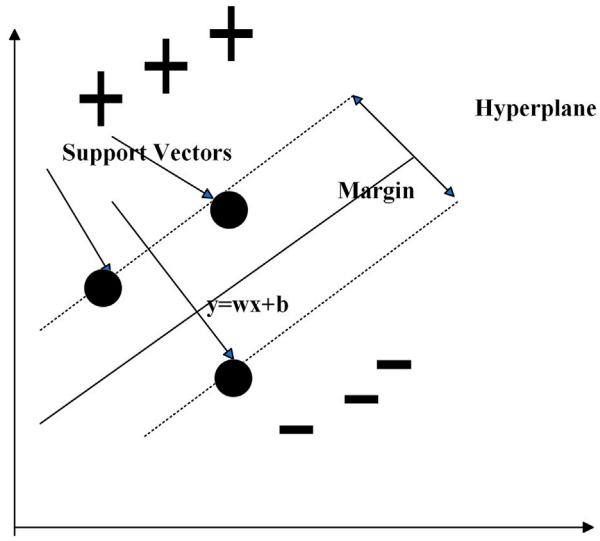


FIGURE 2
Schematic diagram of the support vector machine.

3.2 Machine learning algorithm modeling

Anticipating the load on power grids poses a significant obstacle within the energy industry in the current decade. Machine learning algorithms provide frameworks for analyzing the production, consumption, and distribution of power. With the coupling of supervised and unsupervised learning, machine learning derives thousands of algorithms as single and multiple predictive models for forecasting. The development of machine learning focuses on the accuracy and effectiveness of algorithms. The accuracy of the algorithms varies according to the sampling of energy data and modeling. The fundamental principle behind machine learning algorithms involves selecting past power load data as training samples, creating an appropriate network structure, and employing learning algorithms to predict the energy needs within the power sector. Figure 1 illustrates the process of applying machine learning models to power load data.

The training phase involves a cross-validation approach to ensure the efficacy of the applied models. This method is used to ensure that each split produces results independent of the training, thereby minimizing overfitting in the modeling. Consequently, the training process occurs in one partition, comprising 80% of the training data, while the model's performance is assessed in another partition, encompassing the remaining training data. This iterative process involves alternately positioning the validation subset throughout the training dataset.

Moreover, the training process utilizes mini-batch gradient descent to prevent stagnation at local minima and enhance the model's convergence [47]. Upon completion of the training phase with these techniques, the model undergoes evaluation on the test

dataset. This section examines the capabilities and assesses the suitability of various ML and DL models for electricity load forecasting.

The fundamentals of ML models are presented, commencing with SVM, followed by recurrent neural networks (RNNs) and LSTM. Subsequently, the distinctive features of the DL model and ensemble classifier are introduced.

The various algorithms applied in this study are described as follows.

3.2.1 Support vector machine

The SVM learning algorithm was formulated in 1990 and has since found extensive application in forecasting and pattern recognition tasks. Schematic Diagram of Support Vector Machine is shown in Figure 2. SVM learns to estimate input data on a regression line with a designated threshold. In this model, the ideal trend line that best fits the data is referred to as the hyperplane, while the boundary lines delineate the threshold. SVM maps training data using mathematical equations, called kernels, to determine the hyperplane containing the maximum input data within the boundary lines. It can exhibit linear, nonlinear, or sigmoid characteristics. In nonlinear SVM, the feature data undergo mapping from one plane to another, and data points are segregated in a nonlinear manner, with the decision factor determined by the margin function of the support vector.

$\frac{1}{|w|}$ is the margin. The concept of SVM revolves around maximizing the margin, as given in Eq. 1

$$\text{Max}\left(\frac{1}{\|w\|}\right). \tag{1}$$

This concept can be expressed as shown below in Eq. 2

$$\text{Min}\,\|w\|. \tag{2}$$

The hyperplane for the equation is acquired in the following manner, as given in Eq. 3 (Jiang et al., 2020; Satre-Meloy et al., 2020; Wang R. et al., 2020; Ibrahim et al., 2020):

$$\begin{aligned} w.x_i + b \geq 1, \text{if } y_i = +1, \\ w.x_i + b \leq -1, \text{if } y_i = -1. \end{aligned} \tag{3}$$

Here, w is the weight vector, x is the input vector, and b is the bias.

The formula for minimizing the support vector is as follows and given in Eqs 4, 5:

$$\min \frac{\|w\|^2}{2} + C.\frac{1}{n}\sum_{i=1}^{n}\xi_i, \tag{4}$$

$$y_i\,(w.x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \tag{5}$$

where ξ_i is some units of distance away from the correct hyperplane in the incorrect direction.

"C" is the hyperparameter, which is always a positive value. If C increases, the acceptance of out-of-bound values also escalates. Conversely, as C approaches 0, tolerance diminishes, thereby simplifying the issue and, hence, neglecting the impact of slack.

SVM emerges as a robust machine learning algorithm with promising benefits for electricity load forecasting, owing to several key factors.

- Nonlinearity: SVM adeptly captures complex, nonlinear relationships between electricity consumption and relevant features.
- Outlier resilience: Its design provides inherent robustness against outliers, ensuring stable performance even in the presence of anomalous data points.
- Flexibility: SVM offers adjustable parameters that enable fine-tuning to strike a balance between model complexity and generalization, thereby enhancing adaptability to fluctuations in electricity data.
- Handling high-dimensional data: SVM demonstrates efficacy in handling datasets with a large number of features without compromising its predictive performance.

3.2.2 Recurrent neural network

The RNN is a specialized form of the artificial neural network specifically crafted to analyze sequential time-series data. A key advantage of RNNs lies in their capability to process signals in both forward and backward directions. This is possible by creating network loops and allowing internal connections between hidden components. Due to their internal connections, RNNs are especially adept at utilizing information from preceding data to anticipate future data. Moreover, RNNs enable the exploration of temporal correlations among different datasets (Wang et al., 2019; Wang et al., 2020; H. Ghoddusi et al., 2019). Figure 3 shows the processing of the RNN.

The input to an RNN cell at time step t is typically symbolized as x_t, and the hidden state at time step t is designated as h_t. The output at time step t is denoted as y_t. The cell is also equipped with parameters, such as weights and biases, denoted as W and b, respectively.

Every concealed layer operates based on two inputs: X_t and H_{t-1}. The output Y_t is influenced by the input from the hidden layer (h_t) at time t. These two functions are articulated in Eqs 6, 7 as follows:

$$h_t = f_h(x_t, h_{t-1}), \tag{6}$$

$$\hat{y_t} = f_o(h_t). \tag{7}$$

The input, output, and concealed state of an RNN cell are typically computed using the subsequent Eqs 8, 9:

$$h_t = \phi_h\,(W_{xh}\,.\,x_t + W_{hh}\,.\,h_{t-1} + b_h), \tag{8}$$

$$\hat{y_t} = \phi_o\left(W_{yh}\,.\,h_t + b_y\right). \tag{9}$$

In the transition hidden-layer function of $\varnothing_h$, a nonlinear activation function, like a sigmoid or tanh function, is typically incorporated.

The function $\varnothing_o$ is a nonlinear activation function, such as a rectified linear unit (ReLU), and is derived by computing the dot product of the output weight with the hidden layer h_t and then adding the bias term.

RNNs present numerous advantages for electricity demand forecasting, including the following:

1. Adaptability: RNNs can accommodate variable-length input sequences, making them flexible in handling datasets with missing data.

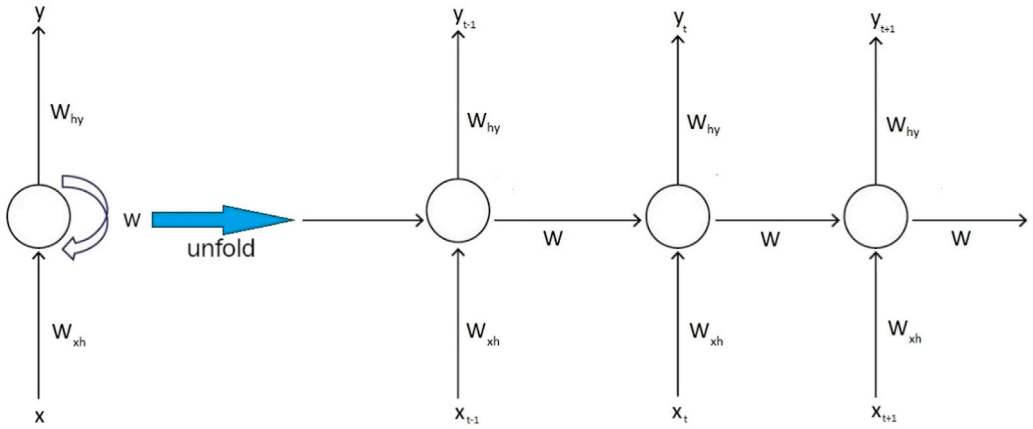


FIGURE 3
Process diagram of the recurrent neural network.

2. Resilience: RNNs demonstrate robustness against noise and outliers present in the data. They possess the capability to detect and filter out irrelevant patterns, ensuring reliable forecasting outcomes.
3. Global context awareness: RNNs capture global information and dependencies across various time scales, allowing for a comprehensive understanding of the underlying patterns influencing electricity consumption.
4. Interpretability: RNN architectures facilitate better comprehension of the patterns contributing to electricity consumption, enhancing the interpretability of forecasting models.

3.2.3 Long short-term memory

The LSTM architecture was initially proposed by Schmidhuber in 1997 (Wang R. et al., 2020). Since its inception, the LSTM architecture has undergone subsequent developments by different researchers (Xue et al., 2019). LSTM was originally devised to combat the issue of vanishing gradients encountered in typical recurrent neural networks when handling long-term dependencies. Unlike a regular RNN, the hidden layers of LSTM possess a more intricate structure, consisting of a sequence of recurring modules. Each hidden layer within LSTM incorporates gate and memory cell concepts. The memory block includes four key elements: an entry gate, an exit gate, a self-connected memory cell, and a deletion gate. The entry gate regulates the activation of the memory cell, while the exit gate controls when to transfer information to the next network layers. Meanwhile, the deletion gate aids in discarding previous input data and resetting the memory cells. Additionally, multiplicative gates are strategically utilized to enable memory cells to retain information over extended periods. This specific architectural design significantly mitigates the vanishing gradient problem encountered in traditional RNNs (Hu et al., 2020; Qiao et al., 2020).

The LSTM features an input x(t), which can originate from either the output of a CNN or the input sequence directly. h_{t-1} and c_{t-1} are the inputs from the LSTM of the previous time step. o_t is the output of the LSTM for this time step. The LSTM also produces c_t and h_t for use by the LSTM in the next time step.

The forget gate f_t, as provided in Eq. 10, decides which previously stored information to maintain or discard upon receiving new data:

$$f_t = \sigma_g\left(W_f \,.\, x_t + U_f \,.\, h_{t-1} + b_f\right), \tag{10}$$

where σ_g is the activation function. The sigmoid activation function is commonly utilized because it condenses information within the interval [0, 1]. This allows the gate to determine the importance of information, for example, whether the value is close to or equals 1, indicating significance, or close to or equals 0, implying insignificance.

The input gate decides which fresh information to retain in the cell state. Initially, Eq. 11 determines what information needs updating i_t.

$$i_t = \sigma_g\left(W_i \,.\, x_t + U_i \,.\, h_{t-1} + b_i\right). \tag{11}$$

Via Eq. 12, the sequence, $c_t^{'}$, is regulated. Like the forget gate, the sigmoid activation function is often employed to retain essential information.

$$c_t^{'} = \sigma_c\left(W_c \,.\, x_t + U_c \,.\, h_{t-1} + b_c\right). \tag{12}$$

By amalgamating the outcomes of the aforementioned gates with the previously retained information in the cell state, c_{t-1}, the value of the cell state denoted by c_t, as provided in Eq. 13, undergoes modification.

$$c_t = f_t \, c_{t-1} + i_t \, c_t^{'}. \tag{13}$$

Ultimately, the output gate dictates the output of the neuron. This result integrates the previously stored information with the fresh data and details obtained from the cell state, as shown in Eq. 14:

$$o_t = \sigma_g\left(W_o \,.\, x_t + U_o \, h_{t-1} + b_o\right), \tag{14}$$

$$h_t = o_t \, \sigma_c\left(c_t\right). \tag{15}$$

The activation function governing the cell state σ_c is described in Eq. 15. The tanh activation function is used to allocate weights to these maintained values.

LSTM networks stand out for their ability to excel in scenarios reliant on temporal data, making them particularly advantageous for electricity load forecasting. The key advantages of LSTM include

1. Temporal modeling prowess: LSTM architectures are specifically tailored to model temporal dependencies in data, making them well-suited for time-series prediction tasks.
2. Context preservation: With their long short-term memory, LSTM networks can effectively capture short-term dependencies while accommodating irregular or missing data points. Additionally, they possess the capability to grasp long-term dependencies within the data.
3. Scalability: LSTM networks demonstrate effectiveness in capturing complex patterns from extensive historical data, enabling robust forecasting in scenarios with diverse and extensive datasets.
4. Robustness: LSTM architectures exhibit resilience against noise and outliers present in the data, ensuring reliable forecasting outcomes even in the presence of data irregularities.

3.2.4 Deep learning

The potential impact of deep learning on power load forecasting methods is substantial. Deep learning is a sub-field of machine learning, which is based on artificial neural networks. In the energy domain, the precision of power consumption prediction is significantly impacted by the processing of forecasted data, making deep neural networks highly relevant. In general, neural networks, better known as MLP, also referred to as feed-forward artificial neural networks, consist of multiple layers that establish connections between the input and output (Ahmed et al., 2020). Each layer comprises its own set of neurons. The number of input neurons typically aligns with the number of features, while the number of output neurons corresponds to the variables to be predicted. The quantity of hidden neurons and layers varies based on the specific problem. As the number of hidden layers or neurons increases, the model's ability to extract complex patterns improves, albeit at the expense of heightened complexity. In such scenarios, the internal operations of the neurons rely on the activation of preceding neurons.

This progression extends from the input layer to the output layer through the neural interconnections. Moreover, if a node possesses multiple inputs, the final value of its function is the sum of the individual values of its functions and their connections. Each iteration of the training process concludes with backward propagation, where the error is disseminated back to the input layers and the weights are adjusted.

In many deep learning problems, we aim to predict an output z using a set of variables X. In this scenario, we assume that for each row of the database Xi, there exists a corresponding prediction z, as given in Eqs 16, 17:

$$z = \sum w_i x_i + b_i, \tag{16}$$

$$a = \psi\,(z), \tag{17}$$

where b_i is the bias. W_i is the weight. ψ is the activation function. "a" is the final output.

Deep learning emerges as a widely employed model, highly conducive to electricity consumption forecasting, owing to several notable benefits:

1. Nonlinear modeling: MLP exhibits prowess in capturing and modeling intricate nonlinear relationships within data.
2. Versatility and customization: With an extensive array of hyperparameters, MLP offers significant flexibility in configuring network architecture and selecting activation functions, tailoring the model to specific forecasting requirements.
3. Robustness to missing data: MLP demonstrates effectiveness in handling missing data, ensuring smooth operation even in datasets with incomplete information.
4. Adaptability to evolving patterns: Equipped with the capability to learn and adjust to dynamic changes in complex data patterns, MLP showcases resilience in forecasting scenarios characterized by evolving trends and behaviors.

3.2.5 Ensemble classifier

A practical approach to enhancing load forecasting accuracy is using an ensemble learning strategy based on artificial neural networks. This ensemble consists of two essential components: a technique for generating sub-samples from the training set and a method for combining them (Ahmad et al., 2020a).

It is crucial to implement both strategies to enhance the overall performance. To form the ensemble, bagging uses a technique that generates ANN models. This is achieved by training them individually on distinct training designs by generating bootstrap replicas of the original training data. In contrast, boosting involves gradually learning ANN models. Using bagging and boosting has yielded positive results in overcoming load forecasting challenges. Nonetheless, we propose a synergistic approach that combines bagging and boosting to maximize their unique capabilities in minimizing variance and bias (Ahmed et al., 2020; Antonopoulos et al., 2020). Therefore, through training conducted with bagging, improvements in generalization are achieved by reducing the model's sensitivity to data variations.

Ensemble learning stands out as a favored ML approach for electricity consumption forecasting due to several compelling factors, highlighted as follows:

1. Regulation and management: Ensemble learning provides a diverse set of regularization methods to manage data complexity effectively, curbing overfitting tendencies and bolstering generalization capabilities.
2. Nonlinearity detection: It adeptly discerns and incorporates nonlinear associations within electricity consumption patterns, enabling the modeling of intricate relationships.
3. Adaptability and efficacy: With its adeptness in handling extensive datasets characterized by high-dimensional feature spaces, ensemble learning demonstrates scalability and operational efficiency.
4. Versatility in optimization: Boasting a broad spectrum of hyperparameters, ensemble learning offers flexibility in fine-tuning model settings to enhance predictive performance and adapt to diverse forecasting scenarios.

3.3 Non-linear complexity handling in machine learning

SVM: SVMs tackle nonlinear complexities by mapping input data into a higher-dimensional space, where nonlinear relationships

can be discerned through kernel functions. However, SVMs may encounter challenges with extremely large datasets and can incur high computational costs.

LSTM networks: LSTMs employ nonlinear activation functions such as the sigmoid and hyperbolic tangent (tanh) functions. These activation functions introduce nonlinearity within the network, enabling it to grasp intricate patterns and correlations within the data. These architectures can capture temporal dependencies in data, allowing them to handle nonlinear sequences of variable length.

Ensemble classifiers: Ensemble methods amalgamate multiple base learners to enhance predictive accuracy. They confront nonlinear complexities by consolidating predictions from diverse models. Techniques like bagging, boosting, and stacking effectively capture nonlinear relationships in power demand time-series data by leveraging the strengths of various base classifiers.

RNNs: RNNs, including LSTM networks, specialize in capturing temporal dependencies within sequential data. They handle nonlinear complexities by sequentially processing input sequences while retaining an internal state that encapsulates historical information. RNNs are particularly suited for modeling nonlinear dynamics and complex patterns in time-series data. RNNs incorporate nonlinear activation functions such as sigmoid, tanh, or ReLU. These functions introduce nonlinearity into the network, enabling it to capture complex patterns and relationships within sequential data. Deep RNNs can capture increasingly complex patterns and dependencies by hierarchically composing nonlinear transformations.

Deep learning methods: Diverse deep learning architectures, such as convolutional neural networks (CNNs) and autoencoders, offer additional approaches for addressing nonlinear complexities in power demand time-series data. CNNs excel at capturing spatial patterns in multidimensional data, whereas autoencoders learn compact representations of input data for nonlinear feature extraction and prediction. Techniques for regularization, like dropout and L2 regularization, are utilized to curb overfitting, a phenomenon where a model mistakenly learns noise in the training data as a genuine signal. By mitigating overfitting, these methods enhance the model's ability to generalize to unseen data and manage nonlinear complexities more adeptly (ur Rehman Khan et al., 2023).

In summary, SVMs, LSTM networks, ensemble classifiers, RNNs, and other deep learning methods possess unique strengths in managing nonlinear complexities in power demand time-series data. The selection of an algorithm hinges on factors such as data characteristics, pattern complexity, and computational resources.

3.4 Limitations or constraints of various machine learning algorithms for real-world forecasting

LSTM networks require a large amount of historical data to effectively capture long-term dependencies, which may not always be available or reliable in power load forecasting applications. Additionally, training LSTM models involves tuning multiple hyperparameters and architecting complex neural network structures, which can be time-consuming and computationally expensive. Moreover, LSTMs are susceptible to overfitting, especially when trained on noisy or limited datasets, which can lead to poor generalization performance on unseen data (Ahmed et al., 2023).

RNNs encounter vanishing and exploding gradient problems, which can make it challenging to learn long-term dependencies in power load data sequences. Standard RNN architectures have limited short-term memory, which may restrict their ability to capture complex temporal patterns in power load data. Training RNNs can be unstable, particularly when dealing with long sequences or noisy data, as it requires careful initialization and regularization to prevent numerical instabilities.

Ensemble learning methods may struggle with the imbalanced datasets commonly encountered in power load forecasting, where certain load patterns are significantly more prevalent than others, leading to biased predictions. The performance of ensemble methods depends on the selection and diversity of base learners, which can be challenging to determine and may require extensive experimentation. Ensembles can be computationally expensive, especially when combining a large number of base learners or using complex algorithms as base models, which may limit their scalability in large-scale power load forecasting tasks.

SVMs provide little insight into the underlying relationships between input features and power load predictions, making it difficult to interpret the model's decisions and identify influential factors. The performance of SVMs is highly dependent on the choice of kernel function, which may require domain expertise and extensive experimentation to identify the most suitable kernel for power load forecasting. SVMs may face scalability issues when applied to large-scale power load forecasting problems, as they require storing support vectors and computing kernel functions for all data points, leading to increased memory and computational requirements.

Deep learning algorithms, including LSTM and RNNs, require large volumes of labeled data to effectively learn complex patterns in power load data, which may not always be available or feasible to acquire. Deep learning models have high model complexity due to their deep architectures and large number of parameters, which can make them prone to overfitting, especially in power load forecasting tasks with limited data. Training deep learning models can be time-consuming, particularly when dealing with large datasets and complex architectures, which may hinder real-time or near-real-time forecasting applications.

3.5 Machine learning models for diverse datasets and time frames

The LSTM machine learning algorithm is particularly well suited for handling diverse datasets and time frames in power load forecasting due to its ability to capture long-term dependencies in sequential data. LSTM can effectively handle various types of data encountered in power load forecasting, including historical load data, weather variables, time of day, day of the week, and holiday indicators. It can process multivariate time series data, incorporating multiple features to make accurate load predictions. LSTM is versatile enough to forecast power load at different time resolutions, ranging from hourly to daily, weekly, or even monthly predictions. It can capture both short-term fluctuations

and long-term trends in energy consumption patterns, making it adaptable to different forecasting horizons (Dinh et al., 2018).

RNNs are well suited for handling diverse datasets encountered in power load forecasting, including historical load data, weather variables, time-related features, and other relevant factors. They can effectively process multivariate time-series data, incorporating multiple input features to make accurate load predictions. RNNs can be applied to forecast power load at various time resolutions, ranging from short-term (e.g., hourly) to long-term (e.g., monthly or yearly) predictions. They can capture both short-term fluctuations and long-term trends in energy consumption patterns, making them adaptable to different forecasting horizons. RNNs are specifically designed to capture temporal dependencies in sequential data. They can learn from the sequential nature of time-series data, identifying patterns and relationships between past, present, and future load values. This makes them effective in capturing seasonality, trends, and periodic fluctuations in power load data.

SVMs are versatile classifiers that can handle diverse datasets encountered in power load forecasting, including historical load data, weather variables, time-related features, and other relevant factors. They can effectively model complex relationships between input features and output labels, making them suitable for multivariate time series data. SVMs can be applied to forecast power loads at various time resolutions, ranging from short-term (e.g., hourly) to long-term (e.g., monthly or yearly) predictions. They can capture both linear and nonlinear relationships in energy consumption patterns, making them adaptable to different forecasting horizons.

Ensemble learning algorithms can handle diverse datasets encountered in power load forecasting, including historical load data, weather variables, time-related features, and other relevant factors. By combining multiple base models, ensemble methods can leverage the strengths of different algorithms to improve prediction accuracy and robustness across various datasets. Ensemble learning techniques can be applied to forecast power loads at different time resolutions, ranging from short-term (e.g., hourly) to long-term (e.g., monthly or yearly) predictions. They can combine forecasts from multiple models trained on different time frames to generate more accurate predictions that capture both short-term fluctuations and long-term trends in energy consumption patterns. Ensemble learning algorithms use various combination strategies, such as averaging, stacking, or boosting, to integrate predictions from multiple base models. These combination strategies can adapt to different forecasting scenarios and data characteristics, ensuring optimal performance across diverse datasets and time frames.

Deep learning algorithms, such as CNNs, RNNs, and deep belief networks (DBNs), can handle diverse datasets encountered in power load forecasting. These algorithms are capable of processing various types of data, including time-series data, spatial data, and multi-modal data, making them suitable for analyzing complex relationships in power load data. Deep learning algorithms can be applied to forecast power loads at different time resolutions, ranging from short-term (e.g., hourly) to long-term (e.g., monthly or yearly) predictions. RNNs, in particular, are well suited for capturing temporal dependencies in time-series data, allowing them to generate accurate 700 forecasts across different time frames (Chen et al., 2018).

4 Experimental analysis and results

4.1 Case study

The assessment of the efficacy of machine learning algorithms in forecasting electric load demand involves the utilization of MATLAB software version R2018. Machine learning algorithms such as SVMs, RNNs, LSTM, DL, and ensemble learning (EM) are applied for the load data forecast. Comparative analysis is presented in *Results*. The analysis utilizes electricity demand data from UT Chandigarh, India, spanning the last 5 years. This dataset encompasses a wide range of demand patterns, including weekly, monthly, and yearly variations. Additionally, the analysis considers seasonal variations such as summer, rainfall, and winter, assessing data variability based on both average and peak values. To mitigate the consequences of missing data and noise, the data undergo transformations. Any missing attribute data are filled using the average value of the available data. Optimizing algorithm hyperparameters relies on transforming energy consumption data. This assessment of forecasting accuracy relies on parameters such as RMSE, normalized mean squared error (NMSE), MI, and MAE. These parameters are computed for all the algorithms considered. The formulation of these parameters is described as follows (Choi et al., 2020; Lu et al., 2019; Zhang et al., 2021).

The time-related parameters—specifically, hour of the day, day of the week, and day of the year—are derived from the available data. Data standardization is executed to ensure accurate adjustments across all employed models. Features are normalized based on their mean values, while sine and cosine functions are applied to the temporal parameters.

To compare results obtained with different methods, the dataset is divided into training, validation, and testing sets, with proportions of 80%, 10%, and 10%, respectively. Models undergo training via cross-validation on the training set, with accuracy evaluated using the validation set. The test set functions as an autonomous dataset for the ultimate assessment of model adjustments. During training, a patience of 100 epochs, a batch size of 64, and an Adam optimizer with a learning rate of 0.001 are employed.

The SVM model is established with a tolerance of 0.001 and a regularization parameter of 1. The DL model comprises 4 hidden layers with 100, 75, 50, and 25 neurons, respectively. The first three hidden layers are equipped with linear activation functions, while ReLU is utilized for the fourth hidden layer and the output layer.

The RNN architecture features 2 hidden layers with 40 and 20 neurons, respectively. The hidden layer with 40 neurons uses a linear activation function, while ReLU is utilized in the final hidden layer and the output layer.

In LSTM modeling, short-term memory is imposed on six time steps, corresponding to the 6 h prior to the forecasted instant. The initial hidden layer includes an LSTM layer with 64 neurons, followed by 2 additional hidden layers with 40 and 20 neurons, respectively, which serve as dense layers. Linear and ReLU activation functions are applied in these two hidden layers. The LSTM layer is equipped with predefined activation functions on its gates, while the output layer also uses a ReLU activation function.

4.2 Performance metrics

The performance comparison of the models is derived by normalizing the metrics to the peak electricity consumption in the series. Consequently, the models are visualized based on NMSE, RMSE, MAE, and MI.

4.2.1 Normalized mean squared error

The NMSE assesses the mean squared disparity between predicted and actual values, adjusted by the variance of the actual values. It is computed by averaging the squared errors and dividing by the variance of the actual values. The NMSE ranges from 0 to infinity, with lower scores denoting heightened accuracy. The NMSE is less sensitive to outliers compared to the raw mean squared error (MSE). It provides a normalized measure of the error, making it easier to compare across diverse datasets with differing scales and magnitudes.

The values within the predictive model set are represented by $\{r^n_{t+1}, \hat{r}^n_{t+1}\}^N_{n=1}$.

The NMSE is determined in Eq. 18 and is as follows:

$$NMSE = \frac{1}{N}\frac{\sum_{n=1}^{N}\left(r^n_{t+1} - \hat{r}^n_{t+1}\right)^2}{var\left(y\right)}. \tag{18}$$

Here, n is the number of samples or data points. y_i is the actual or observed value of the target variable for the *i*th sample. $\hat{y}_i$ is the predicted value of the target variable for the *i*th sample. Var(y) is the variance of the actual or observed values of the target variable.

4.2.2 Root mean squared error

The RMSE mirrors the NMSE but delivers the square root of the mean squared deviation between the predicted and actual values. It is computed as the square root of the average of squared errors. The RMSE shares the same units as the initial data, simplifying interpretation. Similar to the NMSE, diminished RMSE values signify heightened accuracy, with 0 representing the optimal outcome.

The RMSE is calculated as the square root of the MSE, and it is defined in Eq. 19 as

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(r^n_{t+1} - \hat{r}^n_{t+1}\right)^2}. \tag{19}$$

4.2.3 Mean absolute error

The MAE gauges the mean absolute deviation between the predicted and actual values. It is computed as the average of absolute errors. The MAE demonstrates lower sensitivity to outliers than the RMSE since it refrains from squaring the errors. Diminished MAE values signify heightened accuracy, with 0 representing the optimal outcome.

The MAE is given in Eq. 20 as

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left[r^n_{t+1} - \hat{r}^n_{t+1}\right]. \tag{20}$$

The inequality remains valid for the two metrics: MAE ≤ RMSE.

Both of these error measures are regarded as informative in evaluating the model's performance.

4.2.4 Mutual information

Mutual information (MI) evaluates the level of information captured by the model in contrast to a reference model. It is computed as the disparity in information content between the predicted and actual distributions. Elevated MI values signify enhanced model efficacy in capturing the inherent patterns within the data. MI serves as a tool to appraise the predictive capability of a model relative to more straightforward baseline models.

The measure of dependency between r_{t+1} and u_t is determined in Eq. 21 as

$$MI\left(r_{t+1}; u_t\right) = \sum_{r_{t+1};u_t} p\left(r_{t+1}, u_t\right) log\frac{p\left(r_{t+1}, u_t\right)}{p\left(r_{t+1}\right)P\left(u_t\right)} \approx \frac{1}{N}\sum_{n=1}^{N}\log\frac{p\left(r^n_{t+1}|u^n_t\right)}{p\left(r^n_{t+1}\right)}. \tag{21}$$

MI (r_{t+1}; u_t) = 0, when the two variables are independent.

When the two variables are fully dependent, it is bound to the information entropy, $H\left(r_{t+1}\right) = -\sum_{r_{t+1}} p\left(r_{t+1}\right) log\, p\left(r_{t+1}\right)$.

Based on the earlier assumption, we obtain $r_{t+1}\ u_t \sim N\left(\hat{r_{t+1}}, \beta\right)$.

Under an additional presumption, $r_{t+1} \sim N\left(\mu, \sigma\right)$.

We calculate the parameters β, μ, and σ.

4.3 Result analysis

4.3.1 Weekly load forecast analysis

4.3.1.1 Root mean squared error

Figure 4 shows the variation in the RMSE for weekly load predictions using SVM, EM, RNN, DL, and LSTM methods. The results of this variation are spread throughout the week, with the LSTM model closely resembling the actual load pattern. During periods of significant load fluctuations, SVM and ensemble learning models show larger prediction errors, whereas the LSTM model accurately captures the load trend. In comparison, RNN and DL models exhibit greater prediction deviations than LSTM. The analysis emphasizes that the LSTM model achieves an RMSE metric of 0.13% when used for a load forecast spanning a week. This marks a significant improvement, being 23%, 30%, 46%, and 84% less than the relevant metrics for DL, RNN, SVM, and EM models, respectively.

4.3.1.2 Normalized mean square error

Figure 5 shows the variation in the NMSE for weekly load predictions. As the duration extends, variations in the model's performance become apparent. LSTM consistently exhibits the lowest NMSE, suggesting superior performance among these models over this timeframe. The SVM model showed a notable 33% enhancement from its initial error rate, while the EM model showed an improvement of approximately 20%. This implies a moderate reduction in error, rendering the model marginally more dependable. Notably, the RNN model demonstrated the most significant improvement, with a remarkable 50% reduction in error. This translates to a substantial enhancement in prediction accuracy, positioning the RNN as a preferable choice for load forecasting. Similarly, the DL model witnessed an improvement of approximately 33%, indicating consistent performance stability. This underscores the reliability of LSTM in consistently delivering

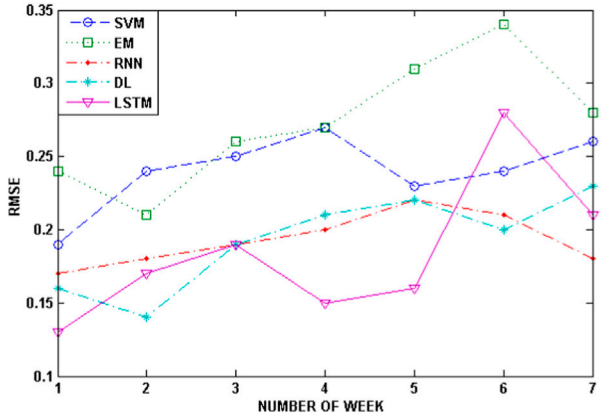


FIGURE 4
RMSE for the weekly load estimations of power load using machine learning algorithms.

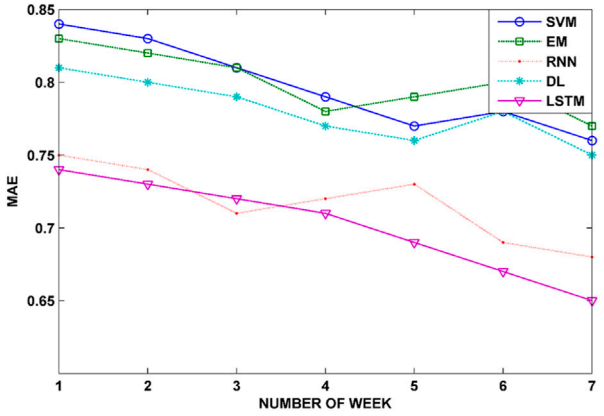


FIGURE 6
MAE for the weekly load estimations of power load using machine learning algorithms.

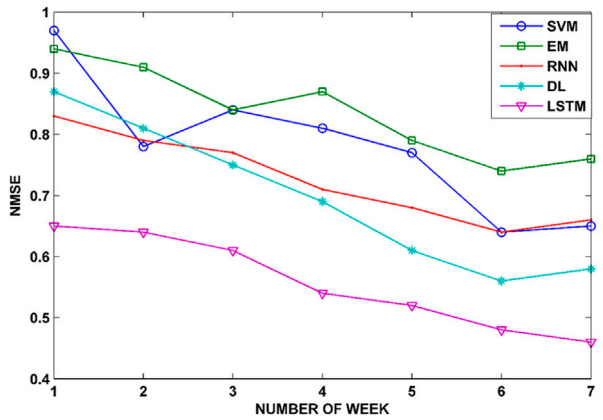


FIGURE 5
NMSE for the weekly load estimations of power load using machine learning algorithms.

accurate predictions. It is worth noting that NMSE values closer to 1 denote poorer performance, thus emphasizing the desirability of lower values.

4.3.1.3 Mean absolute error

The MAE of the LSTM model is measured at 0.74, indicating a substantial reduction compared to the RNN, DL, EM, and SVM models—0.75, 0.81, 0.83, and 0.84, respectively. Figure 6 shows the MAE for the weekly load forecast. The graph represents the performance of different machine learning models (SVM, EM, RNN, DL, and LSTM) over a period of a week.

The SVM model achieved a 5.88% enhancement in prediction accuracy, indicating that its predictions were, on average, 5.88% closer to the actual values. Similarly, the EM model's accuracy improved by approximately 4.88%, resulting in less deviation from the actual values. The RNN model notably enhanced its predictions, reducing the error by 6.10%. Likewise, the DL model showed an improvement of approximately 5.88% in its predictions. The LSTM consistently delivered accurate predictions with minimal error, making it a reliable choice for users seeking stable forecasts.

4.3.1.4 Mutual information

The results of the MI variation demonstrate that higher MI values correspond to better LSTM predictions, whereas SVM and EM exhibit lower MI values, indicating weaknesses compared to the RNN, DL, and LSTM. These findings collectively highlight a significant enhancement in the prediction accuracy of the LSTM method compared to other prediction models.

Over the course of 7 days, the SVM model exhibited a 12.5% improvement in performance. This enhancement translates to a significant increase in accuracy for SVM predictions. Similarly, the EM model showed an improvement of 10.3% in prediction accuracy. The RNN model notably enhanced its predictions by 8.8%. The DL model maintained consistent performance without any decrease. The LSTM consistently delivered reliable predictions with minimal deviation.

4.3.2 Monthly load forecast analysis

Throughout the monthly analysis, we adjusted the hyperparameter settings and conducted the same prediction tests over a 12-month period. Figures 8–11 show the load fitting curves for 12-month load predictions generated by various models. The LSTM model consistently reflects the actual load, while the curves for the SVM and EM models deviate significantly from the actual load, making them the least accurate among all the models. Importantly, the LSTM model proposed in this context shows the closest alignment with the actual load compared to other models.

4.3.2.1 Root mean squared error

Figure 8 shows the RMSE for the monthly load forecast, featuring machine learning algorithms including SVM, RNN, EM, DL, and LSTM. The SVM model's predictions experienced a 7.69% improvement in terms of the RMSE. Similarly, the EM model's accuracy was enhanced by 8.33%. The RNN model notably improved its predictions by 9.09%. Meanwhile, the DL

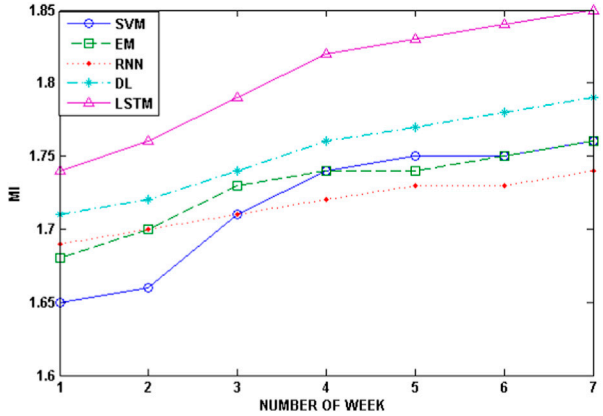


FIGURE 7
MI for the weekly load estimations of power load using machine learning algorithms.

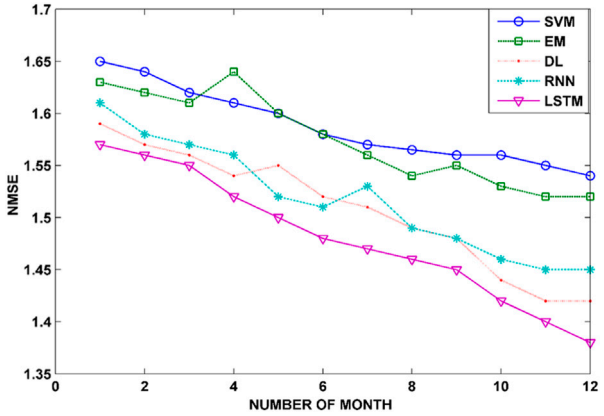


FIGURE 9
NMSE for the monthly power load estimations.

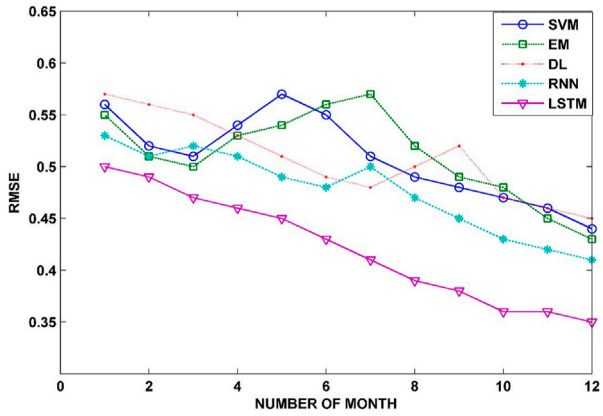


FIGURE 8
RMSE for the monthly power load estimations using machine learning algorithms.

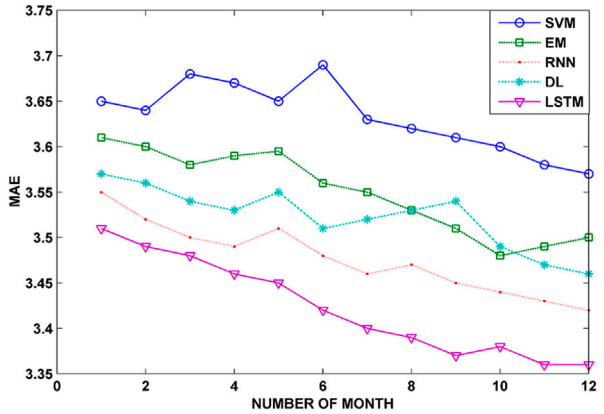


FIGURE 10
MAE for the monthly power load estimations.

model maintained steady performance without any deterioration, demonstrating consistent predictions. The RMSE value remained consistently low for LSTM, indicating its reliability and minimally varying predictions.

4.3.2.2 Normalized mean square error

Figure 9 gives the NMSE for the monthly forecast. Across all months, the LSTM model consistently demonstrates superior performance, exhibiting the lowest NMSE values. This consistency suggests that it is the most accurate model among those compared. Both the SVM and EM models show similar trends, with their NMSE values closely aligned, although SVM slightly outperforms EM. Similarly, the RNN and DL models exhibit comparable trends, with the RNN showing a slight edge over DL.

The SVM model starts with the highest NMSE but shows significant improvement over time, achieving a reduction of approximately 40% in the NMSE. Conversely, the ensemble classifier maintains consistent performance but does not demonstrate as much improvement as SVM, only reducing the NMSE by 15%. Initially, the deep learning model outperforms both EM and SVM models, but its improvement rate slows down, resulting in a reduction of approximately 25% in the NMSE.

The RNN model begins with a lower NMSE and steadily improves over time, achieving a 30% reduction in the NMSE. However, LSTM consistently maintains the best performance throughout, achieving a remarkable 50% reduction in the NMSE. Overall, LSTM emerges as the most consistent and effective model, achieving the highest reduction in the NMSE over the 12-month period.

4.3.2.3 Mean absolute error

In contrast, the RNN and DL models exhibit suboptimal performance across these metrics. Although the RNN model surpasses the DL, EM, and SVM models regarding the MAE, it does not match the performance of the LSTM and DL models in terms of the NMSE.

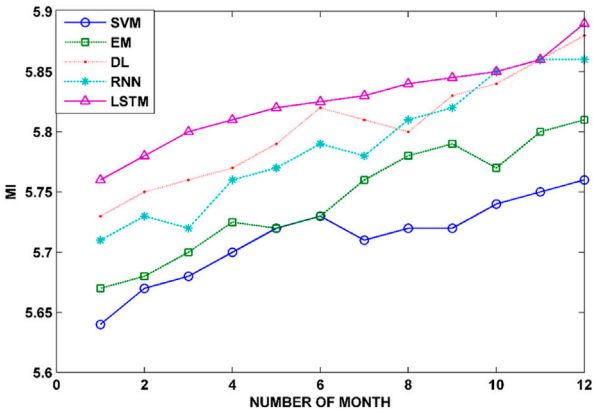


FIGURE 11
MI for the monthly power load estimations.

Consistently, SVM demonstrates the highest MAE across all months, suggesting that it may be the least accurate model in this context. On the contrary, LSTM exhibits a consistent decrease in the MAE from months 0 to 12, indicating an enhancement in accuracy over time. The EM and RNN models show similar patterns, with EM generally performing slightly better than the RNN. DL's performance fluctuates but consistently remains lower than that of SVM and higher than that of LSTM. When considering overall performance, LSTM emerges as the best-performing model, followed by the DL, EM, and RNN models. SVM consistently falls behind the other models in terms of accuracy.

4.3.2.4 Mutual information

Figure 11 shows the mutual information for the monthly load forecast, displaying the performance of five machine learning models (SVM, EM, DL, RNN, and LSTM) across a span of 12 months.

The MI variation results show that higher MI values align with improved LSTM predictions, whereas SVM and EM present lower MI values, indicating relative weaknesses compared to the RNN, DL, and LSTM. The DL and RNN models outperform SVM and EM, showcasing a comparable range of predictions. Overall, these findings underscore a notable enhancement in the prediction accuracy of the LSTM method compared to other models. The RNN exhibits the most substantial improvement (4%), while the SVM experiences a slight decrease (2%).

5 Conclusion and future scope

The vitality of the energy industry hinges on the reliability of power demand forecasting. This study seeks to evaluate the precision of power consumption prediction using machine learning algorithms and hybrid models that incorporate artificial neural networks. Employing these advanced techniques has the potential to significantly improve power load forecasting performance. Several machine learning models are proposed in this study for short-term power load forecasting.

Unlike conventional statistical forecasting models, machine learning algorithms offer numerous advantages, including the effective management of nonlinear complexities and the ability to predict both short-term and long-term dependencies within power load time-series data. This study places emphasis on this crucial aspect. Analysis of the results shows that Figures 4–11 reveal that LSTM outperforms SVM, DL, EM, and RNN regression models in predicting electricity consumption. The approach involves using a sliding time window to convert multidimensional data into a continuous feature map input, harnessing the effectiveness of the CNN in spatial feature extraction. Our methodology is centered on a moving window-based LSTM network, enabling the prediction of demand for specific time intervals.

Following experimental verification analysis, it is concluded that the LSTM model exhibits notable enhancements in the MAE, RMSE, and NMSE compared to alternative time-series prediction models, including SVM, DL, EM, and RNN algorithms. Notably, LSTM outperforms the RNN, DL, EM, and SVM by reductions of 1.35%, 9.45%, 12.16%, and 13.51%, respectively, in predicting the load for the week.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found at: https://www.indiastat.com/chandigarh-state/data/power/power-demand-and-supply.

Author contributions

AJ: conceptualization, data curation, formal analysis, methodology, software, validation, writing–original draft, and writing–review and editing. SG: conceptualization, data curation, formal analysis, methodology, software, supervision, validation, writing–original draft, and 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.

References

Ahmad, T., and Chen, H. (2019a). Deep learning for multi-scale smart energy forecasting. *Energy* 175, 98–112. doi:10.1016/j.energy.2019.03.080

Ahmad, T., and Chen, H. (2019b). Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems. *Sustain. Cities Soc.* 45, 460–473. doi:10.1016/j.scs.2018.12.013

Ahmad, T., and Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustain. Cities Soc.* 54, 102010. doi:10.1016/j.scs.2019.102010

Ahmad, T., Huanxin, C., Zhang, D., and Zhang, H. (2020a). Smart energy forecasting strategy with four machine learning models for climate-sensitive and non-climate sensitive conditions. *Energy* 198, 117283. doi:10.1016/j.energy.2020.117283

Ahmad, T., Zhang, H., and Yan, B. (2020b). A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. *Sustain. Cities Soc.* 55, 102052. doi:10.1016/j.scs.2020.102052

Ahmad, W., Ayub, N., Ali, T., Irfan, M., Awais, M., Shiraz, M., et al. (2020c). Towards short term electricity load forecasting using improved support vector machine and extreme learning machine. *Energies* 13, 2907. doi:10.3390/en13112907

Ahmed, R., Sreeram, V., Mishra, Y., and Arif, M. D. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. *Renew. Sustain. Energy Rev.* 124, 109792. doi:10.1016/j.rser.2020.109792

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., et al. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artif. Intell. Rev.* 56, 13521–13617. doi:10.1007/s10462-023-10466-8

Almaghrebi, A., Aljuheshi, F., Rafaie, M., James, K., and Alahmad, M. (2020). Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods. *Energies* 13, 4231. doi:10.3390/en13164231

Antonopoulos, I., Robu, V., Couraud, B., Kirli, D., Norbu, S., Kiprakis, A., et al. (2020). Artificial intelligence and machine learning approaches to energy demand-side response: a systematic review. *Renew. Sustain. Energy Rev.* 130, 109899. doi:10.1016/j.rser.2020.109899

Aslam, S., Herodotou, H., Mohsin, N. S. M., Javaid, N., Ashraf, S., and Aslam, (2021). A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renew. Sustain. Energy Rev.* 144, 110992. doi:10.1016/j.rser.2021.110992

Bakay, M., Sevgül, Ü., and Ağbulut, (2021). Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. *J. Clean. Prod.* 285, 125324. doi:10.1016/j.jclepro.2020.125324

Bedi, J., and Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Appl. energy* 238, 1312–1326. doi:10.1016/j.apenergy.2019.01.113

Caicedo-Vivas, J. S., and Alfonso-Morales, W. (2023). Short-term load forecasting using an LSTM neural network for a grid operator. *Energies* 16, 7878. doi:10.3390/en16237878

Chammas, M., Makhoul, A., and Demerjian, J. (2019). An efficient data model for energy prediction using wireless sensors. *Comput. Electr. Eng.* 76, 249–257. doi:10.1016/j.compeleceng.2019.04.002

Chapaloglou, S., Nesiadis, A., Iliadis, P., Atsonios, K., Nikolopoulos, N., Grammelis, P., et al. (2019). Smart energy management algorithm for load smoothing and peak shaving based on load forecasting of an island's power system. *Appl. energy* 238, 627–642. doi:10.1016/j.apenergy.2019.01.102

Chen, K., Chen, K., Wang, Q., and He, Z. (2018). Short-term load forecasting with deep residual networks. *IEEE Trans. Smart Grid* 99. doi:10.1109/TSG.2018.2844307

Choi, E., Cho, S., and Kim, D. K. (2020). Power demand forecasting using long short-term memory (LSTM) deep-learning model for monitoring energy sustainability. *Sustainability* 12, 1109. doi:10.3390/su12031109

Dinh, V., Phuong, V. V., Duong, M. Q., Kies, A., Schyska, B. U., and Wu, Y. K. (2018). "A short-term wind power forecasting tool for Vietnamese wind farms and electricity market," in 4th International Conference on Green Technology and Sustainable Development (GTSD), Ho Chi Minh City, Vietnam, 130–135.

El-Hendawi, M., and Wang, Z. (2020). An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting. *Electr. Power Syst. Res.* 182, 106265. doi:10.1016/j.epsr.2020.106265

Fan, Guo-Feng, Y.-H., Guo, J.-M., Zheng, W.-C., and Hong, (2019). Application of the weighted k-nearest neighbor algorithm for short-term load forecasting. *Energies* 12, 916. doi:10.3390/en12050916

Fathi, S., Srinivasan, R., Fenner, A., and Fathi, S. (2020). Machine learning applications in urban building energy performance forecasting: a systematic review. *Renew. Sustain. Energy Rev.* 133, 110287. doi:10.1016/j.rser.2020.110287

Gao, M., Li, J., Hong, F., and Long, D. (2019). Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM. *Energy* 187, 115838. doi:10.1016/j.energy.2019.07.168

Ghoddusi, Hamed, G. G., Creamer, N., and Rafizadeh, (2019). Machine learning in energy economics and finance: a review. *Energy Econ.* 81, 709–727. doi:10.1016/j.eneco.2019.05.006

Hao, Y., Dong, L., Liao, X., Liang, J., Wang, L., and Wang, B. (2019). A novel clustering algorithm based on mathematical morphology for wind power generation prediction. *Renew. energy* 136, 572–585. doi:10.1016/j.renene.2019.01.018

Heydari, A., Nezhad, M. M., Pirshayan, E., Garcia, D. A., Keynia, F., and De Santoli, L. (2020). Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm. *Appl. Energy* 277, 115503. doi:10.1016/j.apenergy.2020.115503

Hou, R., Li, S., Wu, M., Ren, G., Gao, W., Khayatnezhad, M., et al. (2021). Assessing of impact climate parameters on the gap between hydropower supply and electricity demand by RCPs scenarios and optimized ANN by the improved Pathfinder (IPF) algorithm. *Energy* 237, 121621. doi:10.1016/j.energy.2021.121621

Hu, H., Wang, L., Peng, L., and Zeng, Y.-R. (2020). Effective energy consumption forecasting using enhanced bagged echo state network. *Energy* 193, 116778. doi:10.1016/j.energy.2019.116778

Ibrahim, M., Sohail, W., Dong, Q., and Yang, (2020). Machine learning driven smart electric power systems: current trends and new perspectives. *Appl. Energy* 272, 115237. doi:10.1016/j.apenergy.2020.115237

Jiang, P., Li, R., Liu, N., and Gao, Y. (2020). A novel composite electricity demand forecasting framework by data processing and optimized support vector machine. *Appl. Energy* 260, 114243. doi:10.1016/j.apenergy.2019.114243

Jørgensen, K., Lau, H. R., and Shaker, (2020). "Wind power forecasting using machine learning: state of the art, trends and challenges," in 2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, August, 2020, 44–50.

Khan, P., Waqas, Y.-C., Byun, S.-J., Lee, D.-H., Kang, J.-Y., Kang, H.-S., et al. (2020a). Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. *Energies* 13, 4870. doi:10.3390/en13184870

Khan, W., Prince, Y.-C., Byun, S.-J., Lee, N., and Park, (2020b). Machine learning based hybrid system for imputation and efficient energy demand forecasting. *Energies* 13, 2681. doi:10.3390/en13112681

Lai, J.-P., Chang, Y.-M., Chen, C.-H., and Pai, P.-F. (2020). A survey of machine learning models in renewable energy predictions. *Appl. Sci.* 10, 5975. doi:10.3390/app10175975

Li, L.-L., Zhao, X., Tseng, M.-L., and Tan, R. R. (2020). Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *J. Clean. Prod.* 242, 118447. doi:10.1016/j.jclepro.2019.118447

Lu, R., Hong, S. H., and Yu, M. (2019). Demand response for home energy management using reinforcement learning and artificial neural network. *IEEE Trans. Smart Grid* 10, 6629–6639. doi:10.1109/tsg.2019.2909266

Mosavi, A., Salimi, M., Ardabili, S. F., Rabczuk, T., Shamshirband, S., and Varkonyi-Koczy, A. R. (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies* 12, 1301. doi:10.3390/en12071301

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

Pawar, P., Tarunkumar, M., and Vittal K., P. (2020). An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation. *Measurement* 152, 107187. doi:10.1016/j.measurement.2019.107187

Prado, Francisco, M. C., Minutolo, W., and Kristjanpoller, (2020). Forecasting based on an ensemble autoregressive moving average-adaptive neuro-fuzzy inference system-neural network-genetic algorithm framework. *Energy* 197, 117159. doi:10.1016/j.energy.2020.117159

Qiao, W., Lu, H., Zhou, G., Azimi, M., Yang, Q., and Tian, W. (2020). A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer. *J. Clean. Prod.* 244, 118612. doi:10.1016/j.jclepro.2019.118612

Reynolds, J., Ahmad, M. W., Rezgui, Y., and Hippolyte, J.-L. (2019). Operational supply and demand optimisation of a multi-vector district energy system using artificial neural networks and a genetic algorithm. *Appl. energy* 235, 699–713. doi:10.1016/j.apenergy.2018.11.001

Santamouris, M. (2020). Recent progress on urban overheating and heat island research. Integrated assessment of the energy, environmental, vulnerability and health impact. Synergies with the global climate change. *Energy Build.* 207, 109482. doi:10.1016/j.enbuild.2019.109482

Satre-Meloy, A. M., Diakonova, P., and Grünewald, (2020). Cluster analysis and prediction of residential peak demand profiles using occupant activity data. *Appl. Energy* 260, 114246. doi:10.1016/j.apenergy.2019.114246

Shaw, R., Howley, E., and Barrett, E. (2019). An energy efficient anti-correlated virtual machine placement algorithm using resource usage predictions. *Simul. Model. Pract. Theory* 93, 322–342. doi:10.1016/j.simpat.2018.09.019

Skala, R., Elgalhud, M. A. T. A., Grolinger, K., and Mir, S. (2023). Interval load forecasting for individual households in the presence of electric vehicle charging. *Energies* 16, 4093. doi:10.3390/en16104093

Somu, N., Raman, G., R, M., and Ramamritham, K. (2021). A deep learning framework for building energy consumption forecast. *Renew. Sustain. Energy Rev.* 137, 110591. doi:10.1016/j.rser.2020.110591

Su, M., Zhang, Z., Zhu, Y., Zha, D., and Wen, W. (2019). Data driven natural gas spot price prediction models using machine learning methods. *Energies* 12, 1680. doi:10.3390/en12091680

Sun, Y., Haghighat, F., Benjamin, and Fung, (2020). A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy Build.* 221, 110022. doi:10.1016/j.enbuild.2020.110022

Ullah, Z., Al-Turjman, F., Mostarda, L., and Gagliardi, R. (2020). Applications of artificial intelligence and machine learning in smart cities. *Comput. Commun.* 154, 313–323. doi:10.1016/j.comcom.2020.02.069

ur Rehman Khan, S., Hayder, I. A., Habib, M. A., Ahmad, M., Mohsin, S. M., Khan, F. A., et al. (2023). Enhanced machine-learning techniques for medium-term and short-term electric-load forecasting in smart grids. *Energies* 16, 276. doi:10.3390/en16010276

Walker, S., Khan, W., Katic, K., Maassen, W., and Zeiler, W. (2020). Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy Build.* 209, 109705. doi:10.1016/j.enbuild.2019.109705

Wang, H., Lei, Z., Zhang, X., Zhou, B., and Peng, J. (2019). A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* 198, 111799. doi:10.1016/j.enconman.2019.111799

Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., et al. (2020a). Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Convers. Manag.* 214, 112909. doi:10.1016/j.enconman.2020.112909

Wang, R., Lu, S., and Feng, W. (2020b). A novel improved model for building energy consumption prediction based on model integration. *Appl. Energy* 262, 114561. doi:10.1016/j.apenergy.2020.114561

Xue, P., Jiang, Y., Zhou, Z., Chen, X., Fang, X., and Liu, J. (2019). Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms. *Energy* 188, 116085. doi:10.1016/j.energy.2019.116085

Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y., et al. (2021). A review of machine learning in building load prediction. *Appl. Energy* 285, 116452. doi:10.1016/j.apenergy.2021.116452

Zheng, X., Yang, M., Yu, Y., and Wang, C. (2023). Short-term net load forecasting for regions with distributed photovoltaic systems based on feature reconstruction. *Appl. Sci.* 13, 9064. doi:10.3390/app13169064

Zhou, Y., Ravey, A., and Péra, M.-C. (2019). A survey on driving prediction techniques for predictive energy management of plug-in hybrid electric vehicles. *J. Power Sources* 412, 480–495. doi:10.1016/j.jpowsour.2018.11.085

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 13 June 2024
DOI 10.3389/fenrg.2024.1395621

Check for updates

OPEN ACCESS

EDITED BY
Yang Yang,
Nanjing University of Posts and Telecommunications, China

REVIEWED BY
Fei Mei,
Hohai University, China
Zhihao Yang,
Yangzhou University, China
Jianyong Zheng,
Southeast University, China
Tuan Anh Le,
Technical University of Berlin, Germany
Deyagn Yin,
Changzhou University, China

*CORRESPONDENCE
Haiou Cao,
haiou79013@163.com

RECEIVED 04 March 2024
ACCEPTED 21 May 2024
PUBLISHED 13 June 2024

CITATION
Cao H, Zhang Y, Ge Y, Shen J, Tang C, Ren X and Chen H (2024), Intelligent substation virtual circuit verification method combining knowledge graph and deep learning.
Front. Energy Res. 12:1395621.
doi: 10.3389/fenrg.2024.1395621

COPYRIGHT
© 2024 Cao, Zhang, Ge, Shen, Tang, Ren and Chen. 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.

Intelligent substation virtual circuit verification method combining knowledge graph and deep learning

Haiou Cao[1]*, Yue Zhang[2], Yaming Ge[1], Jiaoxiao Shen[3], Changfeng Tang[1], Xuchao Ren[1] and Hengxiang Chen[2]

[1]State Grid Jiangsu Electrical Power Company, Nanjing, China, [2]State Grid Nanjing Power Supply Company, Nanjing, China, [3]State Grid Suzhou Power Supply Company, Suzhou, China

The correctness of the intelligent electronic devices (IEDs) virtual circuit connections in intelligent substations directly affects the stability of the system operation. Existing verification methods suffer from low efficiency in manual verification and lack uniformity in design specifications. Therefore, this paper proposes a virtual circuit automatic verification method that combines knowledge graphs with deep learning. Firstly, this method utilizes expert knowledge and relevant standard specifications to construct a knowledge graph of virtual circuits, integrating knowledge from historical intelligent substation configuration files into the knowledge graph. Then, leveraging multi-head attention mechanisms and Siamese neural networks, it achieves matching between the textual descriptions of virtual terminals and standard virtual terminal descriptions. Additionally, a verification process for the virtual terminal port address string is incorporated. Finally, experimental validation confirms the effectiveness of the proposed method and strategy, further enhancing the accuracy of virtual circuit verification.

KEYWORDS

intelligent substation, virtual circuit, knowledge graph, natural language processing, Chinese short-text matching

1 Introduction

As intelligent substations advance rapidly, the IED within substations has exhibited a notable surge in the variety of types, a marked increase in automation levels, and a continuous strengthening of safety requirements (Song et al., 2016; Huang et al., 2017). The use of optical fibers in smart substations replaces the signal transmission through cables in traditional substations. The correct configuration of the virtual circuit formed by the logical connection of virtual terminals in the configuration file is a prerequisite for the reliable and stable operation of the substation. For a typical 220 kV substation, there can even be thousands of virtual circuits, and manual verification alone is difficult to ensure the accuracy of virtual circuit configuration. In terms of automatic verification, it is affected by non-standard design, resulting in low verification accuracy and poor universality. To illustrate the need for validation and the validation process, consider a simplified example of a bus protection device and a line protection device in an intelligent substation. These devices need to communicate correctly through virtual circuits to ensure system stability. The bus protection device may send a signal indicating a fault condition to the line protection device.

If the virtual circuit connecting these devices is misconfigured, the fault signal may not be received correctly, leading to potential system failures. Validating this virtual circuit involves ensuring that the virtual terminal descriptions and port address data match the expected configurations. These characteristics present both opportunities and challenges for the development of automatic verification of virtual circuits for IED. IED enable communication by adhering to the unified digital standard protocol IEC 61850 (Selim Ustun and M. Suhail Hussain, 2020). Similarly, the testing of IED's virtual circuits is also based on parsing IED files in accordance with the IEC 61850 protocol to perform virtual circuit verification (Zhao et al., 2023).

Currently, in verifying IED in intelligent substations through digital network communication, manual inspection is required to assess the correctness of corresponding virtual circuit circuits, including events, signals, and data, based on the descriptions of secondary equipment virtual terminals (Cui et al., 2018). This process is essential for validating the effectiveness of protection and assessing the accuracy of device parameter settings (Fan et al., 2020). However, due to the lack of a unified standard for virtual terminal descriptions, the verification of virtual circuits often relies on manual inspection. With a multitude of equipment in substations, the manual verification of virtual circuits is characterized by inefficiency, prolonged testing cycles, and susceptibility to variations due to personnel experience and working conditions. This approach is prone to omissions and errors, necessitating substantial effort for error correction following detection.

In addressing this issue, scholars have explored automated verification methods for virtual circuit circuits. Hao et al. (2020) instantiated the configuration of IED virtual circuits through sub-template matching of substation configuration description (SCD) files, offering a novel approach to virtual circuit verification. However, the scalability and generalizability of this method need improvement. Zhang et al. (2015) proposed an expert system based on SCD files to match virtual terminals and perform intelligent substation verification, but the classification is relatively simple, and the process is time-consuming. Some scholars used deep learning and other intelligent algorithms to solve the verification problem in intelligent substations. Oliveira et al. (2021) proposed a deep learning based intelligent substation schedule monitoring method. Chen et al. (2021) proposed the implementation of IED self configuration based on the use of natural language processing technology. Ren et al. (2020) utilized DCNN for text classification of intelligent recorder configuration files to achieve port address mapping. While effective for longer texts, this method has limitations with short texts and does not consider other port address data. Wang et al. (2018) calculated the semantic similarity of virtual terminals using word embedding techniques for virtual circuit matching, demonstrating good matching results. However, their tokenization method for word embedding does not consider the global semantic information of the text, leaving room for improvement. Through the above discussion, it is evident that automated verification schemes for IED virtual circuits primarily rely on two approaches: one utilizing template-based matching that has high accuracy but low generalizability, and the other employing intelligent algorithms or deep learning to classify or calculate similarity in configuration information, offering strong adaptability but it is limited to single-device configurations.

Knowledge graph is essentially a semantic network that includes various semantic connections between different entities (Chen et al., 2020), exhibiting superior interpretability and data storage structural performance (Wang et al., 2017). It has been widely applied in various fields, such as data retrieval, recommendation systems, and knowledge reasoning (Guan et al., 2019). For the power system, knowledge graph can integrate dispersed knowledge within the power system, effectively excavating useful latent rules from massive textual information within the power system (Liu et al., 2023). At the same time, the graph data structure of knowledge graph also provides great convenience for human understanding. Currently, research on knowledge graph in the field of power is still in its nascent stage, with relevant literature mainly focusing on application exploration and macro framework design. Li and Wang (2023) proposed a multi-level, multi-category knowledge graph application framework for assisting decision-making in power grid fault handling and preliminarily elaborated on the key technologies and solution approaches within the framework; Tian et al. (2022) utilized the graph structure of knowledge graph to express textual information and their relationships, extracting the information required to construct knowledge graph from operation and maintenance reports, realizing the automatic construction of knowledge graph, and proposing an automatic retrieval method for equipment operation and maintenance.

Therefore, this study proposes a secondary virtual circuit automatic verification method combining knowledge graph with deep learning. This method integrates the advantages of prior knowledge matching and similarity calculation of virtual terminal information, utilizes knowledge graph for virtual circuit information querying and extraction, and employs an improved Siamese neural network to calculate the similarity of virtual terminal information. Thus, achieving accurate and efficient secondary virtual circuit automatic verification.

To provide a comprehensive understanding of our proposed method, this study is organized as follows. Section 2 introduces the construction method of a secondary virtual circuits knowledge graph. Section 3 provides a detailed explanation of the automatic matching process of virtual terminal information based on the improved Siamese neural network model, along with the automated verification process of virtual circuits. Section 4 presents the experimental results and performance evaluation, demonstrating the effectiveness of the proposed method. Finally, Section 5 discusses the implications of our findings and suggests avenues for future research in this domain.

2 Knowledge graph construction

2.1 Intelligent substation configuration file structure

In the intelligent substation secondary system, a single optical fiber can transmit multiple channels of data and the one-to-one correspondence of data transmission is ensured by using virtual terminals. Virtual terminals are not physical terminals; they are used

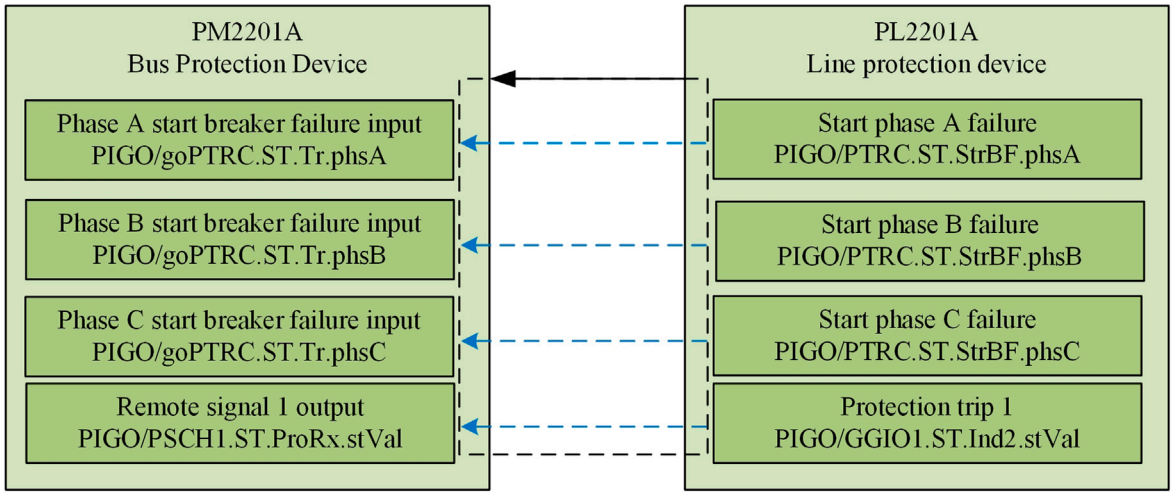


FIGURE 1
Schematic diagram of virtual terminal connections.

to identify loop signals between IEDs and serve as signal connection points during the transmission of generic object-oriented substation event (GOOSE) and sampled value (SV) messages. From the perspective of information transmission, there are two main categories: input virtual terminals and output virtual terminals. From the perspective of message types, there are two main categories: SV virtual terminals and GOOSE virtual terminals.

Taking a certain bus protection device and line protection device as an example, a schematic diagram of virtual terminal connections is illustrated in Figure 1. The bus protection device and line protection device are connected via an optical fiber (solid line with arrow in the diagram), facilitating unidirectional transmission of information in the form of datasets. The virtual terminals are interconnected through virtual connections (dashed line with arrow in the diagram), forming a virtual circuit and thereby achieving a one-to-one correspondence of data.

IEC 61850 is a communication standard among IEDs in intelligent substations that ensures interoperability between different devices. The data modeling technique in IEC 61850 is object-oriented and characterized by a hierarchical tree structure. The order from top to bottom is as follows: physical device, logical device (LD), logical nodes (LNs), data object (DO), and data attribute (DA). Each object in this data structure has a unique data index within the model. The virtual terminal data format specified by the IEC 61850 standard is represented as LD/LN. DO (DA), corresponding to logical device/logical node. data object (data attribute). For the SV virtual terminal, the data attribute (DA) is generally left blank. Therefore, under the IEC 61850 standard, it is ensured that virtual terminals exhibit significant similarities in their data formats.

Additionally, to facilitate interoperability among devices from different manufacturers, designers often include a brief Chinese description on virtual terminals during the equipment design process. This practice aims to aid designers in better understanding and distinguishing virtual terminals associated with devices from various manufacturers.

2.2 Virtual circuit knowledge graph

In an intelligent substation, IEDs encompass a variety of devices, including relay protection devices, merging units, smart terminals, and intelligent recorders. The virtual circuits corresponding to different IEDs are markedly distinct, and virtual circuits of the same type of IED may exhibit certain variations under different states. The verification of virtual circuits involves three aspects:

(1) Precisely determining all the virtual terminals and IEDs essential for configuring the virtual circuit verification.
(2) Clearly defining the hierarchical paths and descriptive features corresponding to different types of IED data formats.
(3) Assessing the correctness of the mapping relationships of virtual terminals based on their distinctive features.

The construction process of knowledge graph involves the following steps:

Step 1: Collect SCD files and other relevant documentation from various intelligent substations. This data provides the raw input needed to build the knowledge graph.

Step 2: Use natural language processing techniques to parse the SCD files and extract relevant information such as IED types, virtual terminal descriptions, and port addresses.

Step 3: Map the extracted information to the ontology. This involves identifying the appropriate entities and relationships in the knowledge graph and ensuring that the extracted data fits into this structure.

Step 4: Use a Siamese neural network with multi-head attention to calculate the semantic similarity between extracted virtual terminal descriptions and the standardized descriptions in the ontology.

Step 5: Integrate the matched data into the knowledge graph, creating links between historical data and standardized models. Manual verification is performed for matches below a predefined threshold to ensure accuracy.

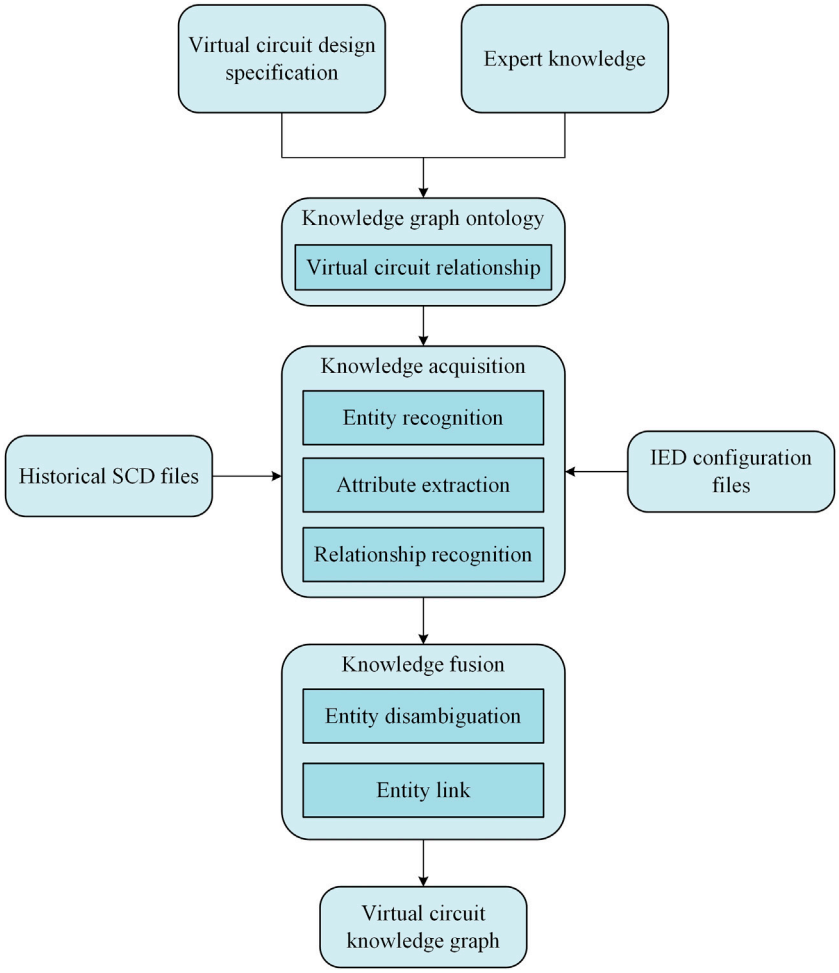


FIGURE 2
Knowledge graph construction method.

Step 6: Continuously update the knowledge graph with new data and validation results to improve its accuracy and comprehensiveness.

This study employed a top–down approach to construct a knowledge graph for the association of virtual circuits. Standardization of virtual terminals for relevant IEDs in intelligent substations was achieved, establishing a standardized model for virtual terminals. Expert knowledge in the context of intelligent substations refers to the domain-specific insights and rules provided by experienced engineers and technicians. This knowledge encompasses the correct configurations of virtual circuits, typical faults, and the standard practices for designing and maintaining these systems. By modeling virtual circuits based on expert knowledge according to IED types and associating them with the standardized virtual terminal data relationships established according to relevant specifications, a knowledge graph ontology was formed. The construction process of knowledge graph is shown in Figure 2. The ontology includes:

(1) Entities: These are the core components such as IEDs, virtual terminals, LNs, DO, and attributes.
(2) Relationships: These define how entities interact with each other. For example, an IED may have multiple virtual terminals, and each terminal can be linked to specific data attributes.
(3) Attributes: These are the properties or characteristics of the entities, such as the type of data transmitted, the logical node identifiers, and port address configurations.

The ontology provides a standardized model that facilitates consistent representation and querying of virtual circuit information.

Utilizing historical data from intelligent substations for knowledge learning, this study parsed historical SCD files of intelligent substations to acquire the associative relationships of IED virtual circuits. Based on the standardized virtual circuit model in the knowledge graph ontology, relationships between standardized virtual circuits and actual virtual circuits within IEDs are established. The constructed virtual circuit knowledge graph is shown in Figure 3. Although there are significant format differences between virtual terminal address configuration data and textual description information, information logic among the same type of IEDs shares strong correlations, and the naming and expressions of identical entities are generally uniform.

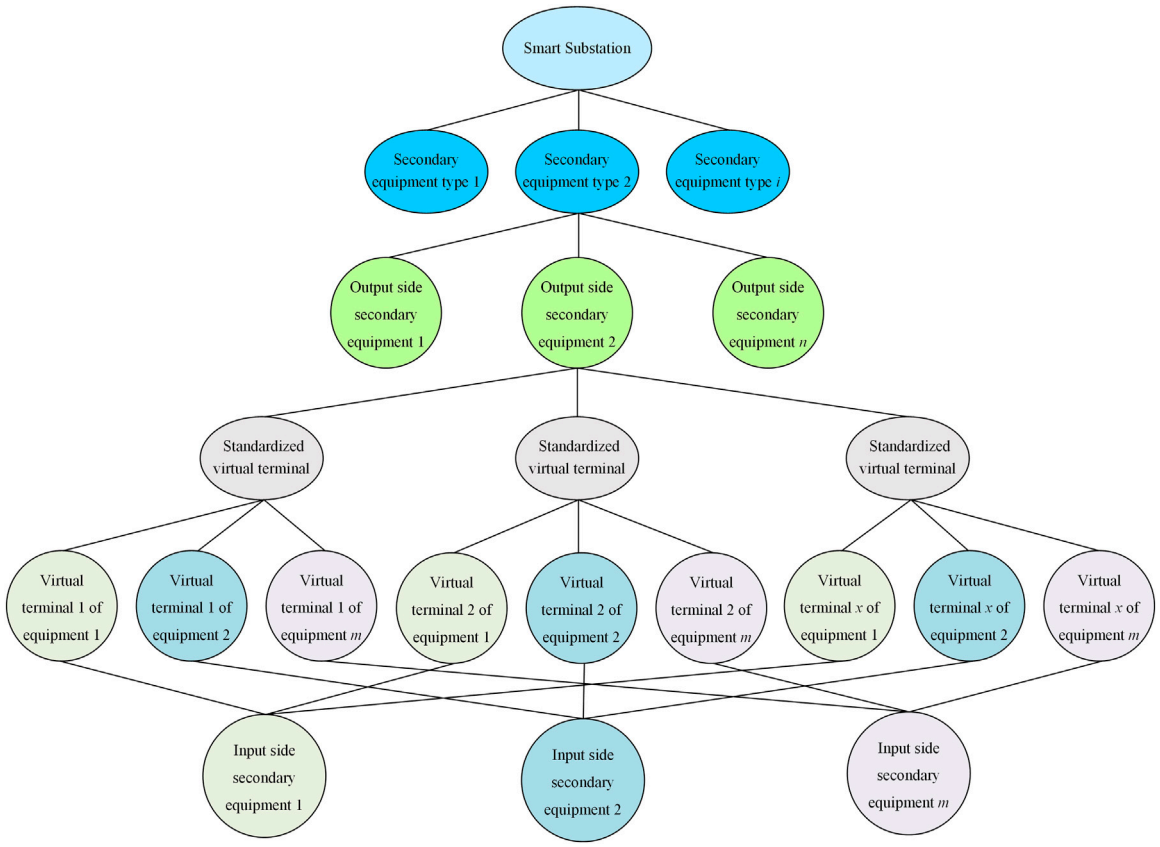


FIGURE 3
Knowledge graph structure diagram.

Using dictionary list data obtained from processing SCD files, three categories of information were extracted: IED device type, virtual terminal address configuration data, and textual description. The entities are linked in triplets: {device type, virtual terminal address configuration data, textual description}. The initial extraction results contained a considerable amount of redundant data. This study removed numerical information from virtual terminal address configuration data, retaining only alphanumeric string information. An index is then constructed by merging IED type and textual description information. The number and types of virtual terminal address configuration data in each index are tallied. Based on the statistical results, the triplets are transformed, retaining only one triplet for each identical virtual terminal address configuration data within the same index.

Establishing a knowledge graph using virtual circuit information from historical SCD files has limitations. Therefore, in this study, based on the similarity calculation between the virtual circuit and the standardized model of virtual circuit data, IED virtual terminals are automatically matched to standardized virtual terminals. Subsequent manual verification is conducted to establish the relationship between IED virtual terminals information and the standard virtual terminals information, achieving knowledge fusion. Knowledge fusion involves integrating information from various sources to create a comprehensive and coherent knowledge graph. In our method, knowledge fusion occurs in two main steps:

(1) Historical Data Integration: We parse historical SCD files to extract virtual circuit configurations from previously implemented substations. This data includes IED types, virtual terminal descriptions, and port address configurations.
(2) Standardization and Matching: The extracted data is compared against the standardized models defined in the ontology. Virtual terminals are matched to their standardized counterparts based on semantic similarity calculations and expert-defined rules. This step ensures that the knowledge graph accurately reflects both historical configurations and standardized practices.

3 Virtual terminal automatic matching

Due to the predominantly Chinese short-text nature of the virtual terminal textual descriptions, a text-matching model is proposed that integrates a multi-head attention mechanism and a Siamese network. The model utilizes a mixed vector of characters and words as input to enhance semantic information, employing a bidirectional gated recurrent unit (Bi-GRU) instead of bidirectional long short-term memory (Bi-LSTM) to reduce parameters and expedite training speed. The multi-head attention mechanism is introduced as a separate module, employing an autoencoding layer to capture semantic features from different perspectives. A Siamese

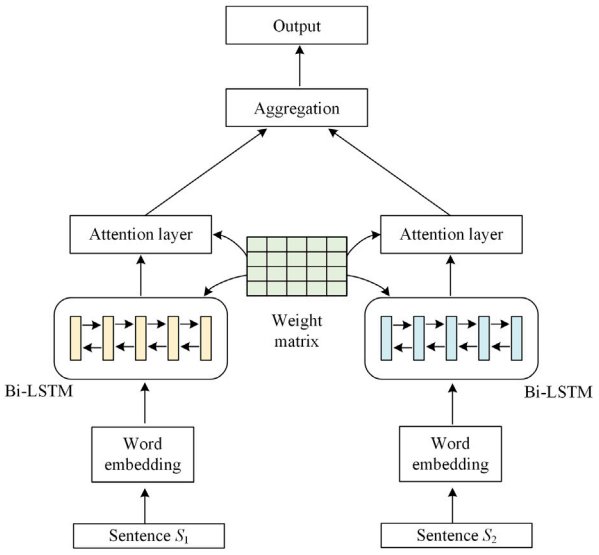


FIGURE 4
Siamese neural network.

network is constructed to transform sentences of varying lengths into sentences of equal length, placing them in the same semantic space. Weight sharing is implemented to reduce half of the training workload. In the interaction layer, the mixed vector of characters and words for one sentence interacts with that of another sentence, utilizing the multi-head attention mechanism to acquire interactive semantic features.

3.1 Siamese neural network model

The schematic diagram of the Siamese neural network model is shown in Figure 4.

The Siamese network model (Liang et al., 2018) is broadly divided into three parts: preprocessing, shared neural network, and information aggregation. The processing flow of the Siamese network is as follows: firstly, preprocess the obtained sentences to obtain word vector representations; then, encode the obtained sentence representations using a Siamese network constructed with Bi-LSTM and attention mechanisms; finally, aggregate information from the sentence representations processed through the Siamese network.

(1) Preprocessing: To begin, sentences containing contextual information are obtained. Each word in the sentence is then represented using word vectors. A sentence is represented as $\bar{V}_{S_1} \in \mathbb{R}^{m\times d}$, $\bar{V}_{S_2} \in \mathbb{R}^{n\times d}$, where m and n represent the lengths of sentences S_1 and S_2 , respectively, and d is the dimensionality of the word vectors.

(2) Shared neural network: After obtaining the representation of sentences, the word vectors pass through a Bi-LSTM algorithm to encode information about the sentences. A standard Bi-LSTM algorithm is employed, representes by Eqs 1–3 as follow:

$$\overrightarrow{h_t} = f\left(U_1 x_t, U_3 \overrightarrow{h_{t-1}}, \overrightarrow{b_t}\right) \tag{1}$$

$$\overleftarrow{h_t} = f\left(U_2 x_t, U_4 \overleftarrow{h_{t+1}}, \overleftarrow{b_t}\right) \tag{2}$$

$$y_t = concat\left(\overrightarrow{h_t}, \overleftarrow{h_t}\right) \tag{3}$$

where, $\overrightarrow{h_t}$, $\overleftarrow{h_t}$, x_t and y_t represent the forward propagation hidden layer state, the backward propagation hidden layer state, the input values of the neurons, and the output values of the hidden layer state at time t, respectively; $\overrightarrow{h_{t-1}}$ represents the forward propagation state at time t-1 and $\overrightarrow{h_{t+1}}$ represents the backward propagation state at time $t+1$; U_1, U_2, U_3 and U_4 denote the weight matrices corresponding to different components, respectively; $\overrightarrow{b_t}$ and $\overleftarrow{b_t}$ represent bias vectors in the forward propagation hidden layer and backward propagation hidden layer, respectively; *concat* denotes concatenate operation.

Following the Bi-LSTM layer, an attention mechanism is introduced. The shared-weight neural network consists of the aforementioned Bi-LSTM layer and an attention mechanism layer.

(3) Information aggregation: The processed representations of the two sentences need to undergo information fusion, and common fusion methods include fully connected neural networks, calculating the cosine similarity, and the Manhattan distance between the two vectors.

3.2 Interactive text matching model integrating multi-head attention mechanism and siamese network

The Siamese network model does not fully leverage interactive information between texts, nor does it adequately capture the representation capabilities of the text. Therefore, this study proposed an improved model, the Interactive Text Matching Model, which integrates a multi-head attention mechanism with a Siamese network, as shown in Figure 5.

Chinese writing is logographic, meaning that each character represents a word or a meaningful part of a word. This differs from alphabetic languages where words are composed of letters. Therefore, the network uses character-based embeddings in addition to word-based embeddings to capture the nuances of Chinese text. This dual representation ensures that both individual characters and their combinations are effectively represented.

This model initially preprocesses two sentences at the input layer, obtaining mixed vectors for both sentences. The resulting mixed vectors undergo normalization through a regularization layer. Subsequently, these processed vectors are separately fed into the difference unit and interaction unit.

In the difference unit, the input texts are first encoded using a bidirectional GRU, constructing a Siamese network. The Manhattan distance is then employed to aggregate the encoded information. Simultaneously, within the interaction unit, the auto-encoding layer is utilized to encode the input sentences separately, forming another Siamese network. Subsequently, the semantic features from the self-encoding layer undergo interaction through the multi-head

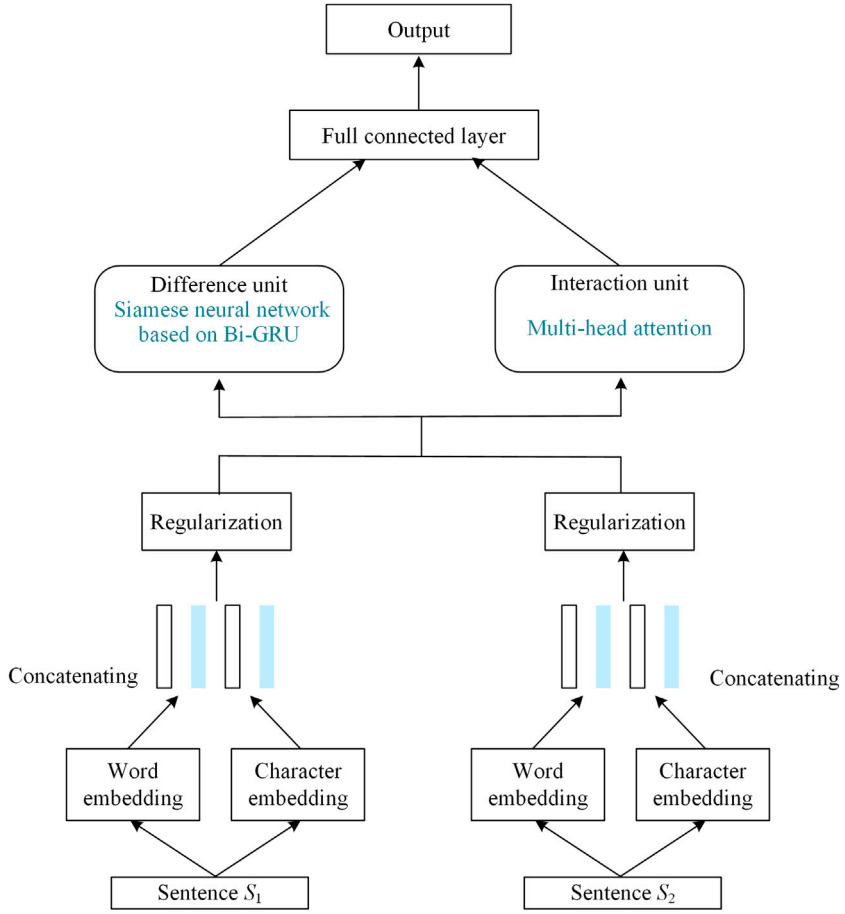


FIGURE 5
Interactive text matching model.

attention mechanism's interaction layer. Finally, the interactive information, along with the information from the difference unit, is output to the fully connected neural network in the output layer. The classification result is obtained through a sigmoid function.

3.2.1 Difference unit

Firstly, the training set is tokenized into word and character level representations using the Jieba segmentation tool. For input sentences S_i, after word and character level segmentation, two representations are obtained: a word sequence and a character sequence. By loading the pre-trained Word2vec model weights, each word or character can be mapped to a vector, extracting the corresponding word/character vectors, the formal description of sentence S_i is given by Eqs 4, 5:

$$S_i = \left\{W_i^1, W_i^2, \ldots, W_i^{lw}\right\} \tag{4}$$

$$S_i = \left\{C_i^1, C_i^2, \ldots, C_i^{lc}\right\} \tag{5}$$

where lw and lc represent the number of words and characters in sentence S_i, respectively.

The length of the word vectors is extended to match the length of the character vectors. Subsequently, concatenating the two representation vectors yields the final hybrid representation vector V_{S_i}, which combines both character and word embeddings. This hybrid word-character vector is then fed into the normalization layer.

$$V_{S_i} = \left\{v_{S_i}^1, v_{S_i}^2, \ldots, v_{S_i}^{N_i}\right\} \tag{6}$$

where N_i represent the length of vector V_{S_i}.

In this study, the Bi-GRU structure was employed to replace the Bi-LSTM algorithm in the Siamese network for text information encoding. A GRU (Cho et al., 2014) is a variant of LSTM with a simplified architecture. It employs an update gate in place of the forget and input gates in LSTM and introduces a new hidden unit. The model structure is simpler than that of LSTM and is represented by Eqs 7–10:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{7}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{8}$$

$$\widetilde{h_t} = \tanh(r_t \cdot U_a h_{t-1} + W_a x_t) \tag{9}$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \widetilde{h_t} \tag{10}$$

where z_t and r_t represent the update gate and reset gate, respectively; $\widetilde{h_t}$ represents the aggregation of the input x_t and the output of the previous hidden layer h_{t-1}; σ denotes the sigmoid function, and tanh denotes the hyperbolic tangent function; W_z, U_z, W_r, U_r, U_a, and W_a are the weight matrices used in training.

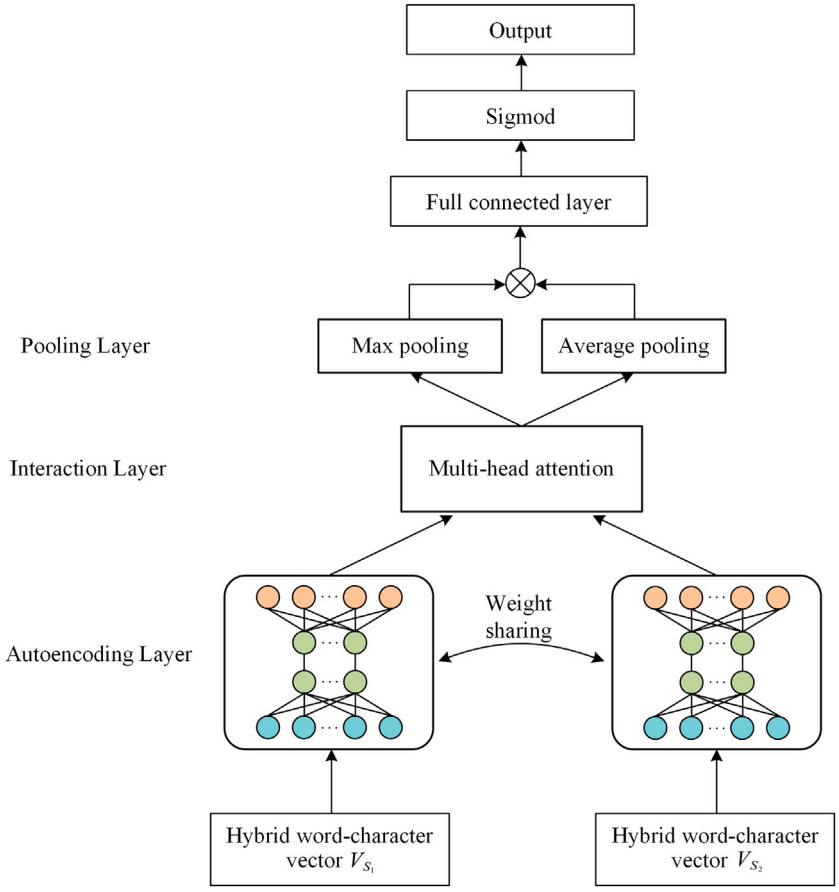


FIGURE 6
Interaction unit structure diagram.

The update gate indicates how much of the current hidden state is inherited by the new hidden state, while the reset gate indicates how much information from the past hidden state should be ignored and reset using the current input. GRU has fewer parameters compared to LSTM, making training faster and requiring less data.

3.2.2 Interaction unit

The Siamese network model introduces the attention mechanism directly after the neural network to extract essential word information from sentences. However, it does not adequately extract the semantic features of sentence pairs and let them interact. Therefore, this study treated the multi-head attention mechanism as a separate unit for computation, extracting interaction features between text pairs. The basic attention network is segmented into different sub-networks, learning more semantic features from various perspectives, with the aim of achieving a comprehensive interaction between semantic features. The interaction unit comprises the autoencoding layer, interaction layer, and pooling layer, as illustrated in Figure 6.

(1) Autoencoding layer

In the autoencoding layer, the mixed-word vectors V_{S_1} and V_{S_2} are separately encoded. This layer consists of two parts, the encoder and the decoder, forming a symmetrical structure. Both the encoder and decoder typically comprise three-layer neural networks, including input, hidden, and output layers, as illustrated in Figure 7. Here, the output of the encoder serves as the input to the decoder, and the outputs of the encoder and decoder are represented by Formulas 11, 12, respectively:

$$Z_i = f_1\left(W_e V_{S_i} + b_e\right) \tag{11}$$

$$V'_{S_i} = f_2\left(W_h \hat{H} + b_h\right) \tag{12}$$

where W_e and W_h represent weight matrices; b_e and b_h are bias vectors; f_1 and f_2 are activation functions; $\hat{H}$ represents the hidden layer vector.

The autoencoder network extracts high-dimensional features through the encoder, reducing the dimensionality to process the output text features, denoted as Z_i. The decoder, employing a symmetric network structure, reconstructs the input V_{S_i} of the encoder to obtain V'_{S_i} aiming to fit an identity function using neural networks and enhance the feature extraction capability of the encoder. The reconstruction process utilizes mean square error as the loss function, and L2 regularization is employed to prevent model overfitting, thereby improving the model's performance on the test set.

(2) Interaction layer

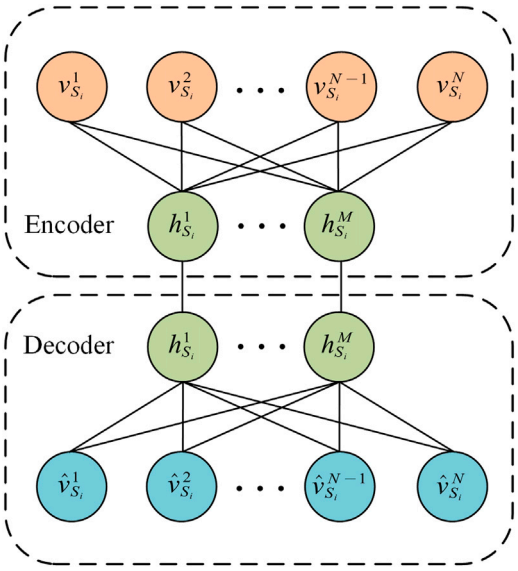


FIGURE 7
Autoencoder structure diagram.

In the encoding layer, each text is individually encoded, obtaining diverse contextual semantic information to enhance the representation of text features. However, due to the Siamese network's inclination to represent semantic feature vectors of different lengths in the same semantic space, the interaction information between texts is overlooked. Compared to typical Chinese text, descriptions of virtual terminals consist of shorter sentences, and there is no contextual semantic relationship between sentences. Therefore, the interaction layer is introduced to separately extract word and sentence interaction features between text pairs. Each word can interact with words in the other text, capturing syntactic and semantic dependencies between text pairs.

This layer is introduced to address the deficiency of the Siamese network in semantic interactions. Multi-head attention mechanisms are employed to extract interaction information between the two texts. The mixed-word vectors after autoencoding $V'_{S_1}$and $V'_{S_2}$are concatenated to form V_{text}, and then utilize multi-head attention based on scaled dot-product attention (Vaswani et al., 2017) to capture interactions between pairs of sentences and can be described by Eq. 13:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}V\right) \tag{13}$$

where Q, K, and V represent the query vector, key vector, and value vector respectively, with d_k being the dimensionality of key vector. In this study, self-attention is employed to extract features of V_{text}, thus Q, K, and V are all equal to V_{text}.

The attention mechanism acts as semantic feature extraction and encoding, providing each word with three vectors. Each operation involves calculating the similarity between a word's query vector and all key vectors through dot-product, resulting in weight coefficients representing the word's relevance to other words. These coefficients are then used to weigh all value vectors to obtain semantic encoding. Multi-head attention allows different attention weights to be assigned to different positions, acquiring better semantic information and effectively preventing overfitting, as describes in Eqs 14, 15 (Guo et al., 2019):

$$Multi-head(Q,K,V) = concat(head_1, head_2, \ldots, head_H)W^O \tag{14}$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^K) \tag{15}$$

where W^O, W_i^Q, W_i^K, W_i^V represent the weight matrix.

(3) Pooling layer

The pooling layer's function is to extract global features from the word vector sequences. This includes both max pooling and average pooling. Each dimension of the word vector reflects different information, and pooling operations help to extract comprehensive information from the word vectors. Considering that max pooling retains prominent information from the word vectors, while average pooling retains information from all word vectors (Bieder et al., 2021), we simultaneously use both max pooling and average pooling, and then concatenate the results, as describes in Eqs 16–18:

$$V_{text,avg} = \sum_{i=1}^{N_{V_{text}}} \frac{V^i_{text,att}}{N_{V_{text}}} \tag{16}$$

$$V_{text,\max} = \max_{i\in[1,N_{V_{text}}]} V^i_{text,att} \tag{17}$$

$$V_{text,pool} = concat\left(V_{text,avg}, V_{text,\max}\right) \tag{18}$$

where $V_{text,avg}$, $V_{text,\max}$, $V_{text,pool}$ represent average pooling result, maximum pooling result, and pooling layer result. $V_{text,att}$ represents the output result of multi-head attention in the interaction layer; $N_{V_{text}}$ represents the length of vector $V_{text,att}$.

For the output, the results X_{itac} and X_{diff} from the interaction unit and the difference unit are hybrid through a fully connected neural network. The final classification result is obtained using the sigmoid activation function, as describes in Eqs 19, 20:

$$R = \text{sigmod}\left(FC\left(X_{itac}, X_{diff}\right)\right) \tag{19}$$

$$FC(X) = \text{relu}(Wx + b) \tag{20}$$

where W represents the weight matrix; b represent bias vector. relu represents the rectified linear unit

The model utilizes mean squared error as the loss function. Additionally, the Adam optimization algorithm is applied to enhance the convergence speed, as describes by Eq. 21 (Reyad et al., 2023).

$$LossMSE = \frac{1}{N}\sum_{1}^{N}\left(Y_i - \hat{Y}_i\right) \tag{21}$$

where N represents the number of samples; Y_i and $\hat{Y}_i$ represent the true value and predicted value, respectively.

3.3 Verification method based on port address data

Through experiments, it was observed that accurately calculating the textual similarity values can reliably detect the desired virtual terminals. However, in some cases, there may be

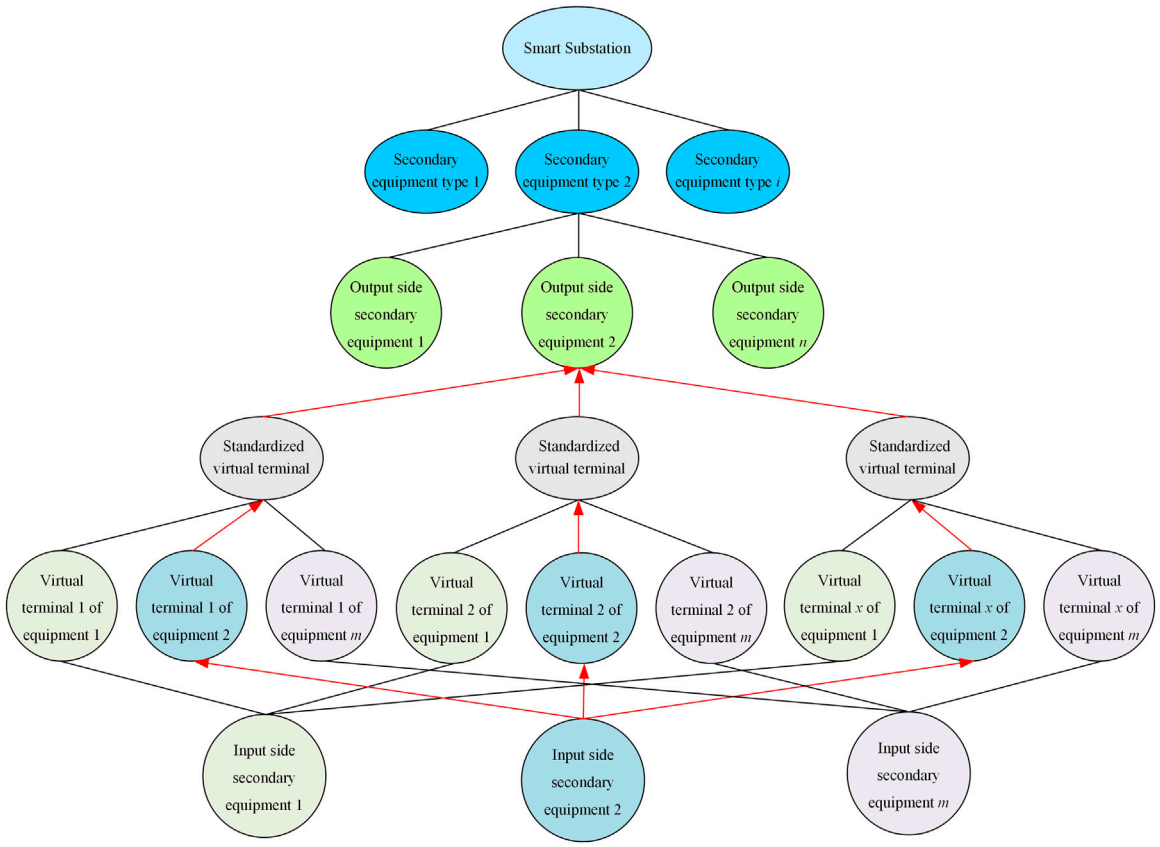


FIGURE 8
Automatic verification path.

more than one matching result. This is attributed to incomplete or the repetitive construction of certain virtual terminal description texts. To distinguish virtual terminal ports in such cases, reliance on port address data information becomes crucial, as each virtual terminal's port address is unique.

Therefore, this study introduced a verification process based on port address configuration data to address situations where textual descriptions alone cannot establish mapping for the data. Leveraging port address configuration data helps discern matching relationships due to the systematic naming patterns inherent in port address configuration data, with most strings carrying practical significance.

The longest common substring (LCS) method proposed by Amir et al. (2020) was employed to assess the similarity of virtual terminal port addresses. The calculation of the virtual terminal port address similarity involves evaluating the similarity of four attributes constituting the virtual terminal port address: logical device, logical node, data object, and data attribute. The weighted average of these individual similarities is taken as the overall similarity of the virtual terminal port address.

3.4 The automatic verification process for virtual circuit

In intelligent substations, the virtual circuit information of various IEDs is integrated into the station-wide configuration SCD file. Before verification, the SCD file is parsed to obtain the IEDs that need to be checked. The output interface addresses and descriptions are extracted from the SCD file to form the output interface information for the corresponding IED.

During the verification process, the knowledge graph provides the reference model against which current virtual circuit configurations are compared. This comparison helps in identifying discrepancies and potential faults. The verification process begins with the IEDs in the knowledge graph. The matching is conducted along the path indicated by the arrows in Figure 8. The device model information parsed from the SCD file is used to search for the corresponding IED in the knowledge graph. The virtual terminal information extracted from the SCD file for the identified IED are then matched with the virtual terminal information in the knowledge graph, specifically matching them to standard virtual terminal information. This process enables the detection of all virtual circuits for the specified IED, achieving automatic verification of virtual circuits in intelligent substations.

If the corresponding IED cannot be matched in the knowledge graph during verification, a search is conducted based on the device type to find the standard output interface addresses corresponding to that device type in the knowledge graph. Subsequently, the similarity between the virtual terminal information of the IED and the standard virtual terminal information is calculated. Any matches below a predefined threshold undergo manual confirmation to establish the association between the IED's

TABLE 1 Partial data samples.

Order	Input virtual terminal		Output virtual terminal	
	Description	Address	Description	Address
1	A-phase voltage sampling value	MU/UATVTR1. Vol1	Voltage A-phase measurement value	SVLDO1/SVINUATVTR1. Vol2. Inst mag. I
2	B-phase voltage sampling value	MU/UBTVTR1. Vol1	Voltage B-phase measurement value	SVLDO1/SVINUBTVTR1. Vol2. Inst mag. I
3	C-phase voltage sampling value	MU/UCTVTR1. Vol1	Voltage C-phase measurement value	SVLDO1/SVINUCTVTR1. Vol2.inst mag. I
4	Merge unit optical port transmission optical power	PIGO/GOINGGIO17. AnIn10. mag. f	Bus merging unit board 1 optical port 1 optical power	MUGO/SCLI9. LigIntes. mag. f
5	Intelligent terminal network port 1 abnormality	PIGO/GOINGGIO 1. SPCSO6. stVal	Bus intelligent terminal network port 1 is abnormal	RPIT/GOAlmGGI O1. Alm6. stVal

virtual terminal addresses and the standard virtual terminal addresses. This association is then stored in the knowledge graph to facilitate knowledge updates.

4 Case study

4.1 Dataset

To validate the effectiveness of the proposed configuration method, a sample set was selected comprising 10 SCD files from 220 kV intelligent substations and 20 SCD files from 110 kV intelligent substations. These samples encompassed 3,869 IEDs and 29,110 records of configured virtual circuits. To conduct experimental analysis, the dataset was divided into training, testing, and validation sets in a ratio of 7:2:1. Since virtual circuit configurations vary significantly across intelligent substations with different voltage levels, data sets were randomly extracted from SCD files of different voltage levels to ensure the generalizability of the proposed method. A subset of the data samples is presented in Table 1. It is evident that the textual descriptions of virtual circuit endpoints share significant similarities, yet there are distinct differences in the address data. In the virtual terminal address data, most strings carry meaningful information; for instance, “MU” signifies a merging unit, and “UATATR” denotes voltage sampling. However, even for the same voltage sampling virtual terminal, the configuration data for virtual terminal addresses can be entirely different. This discrepancy arises due to varying naming conventions among different manufacturers’ IEDs, and certain strings such as “mag” and “AnIn” pose challenges in determining their actual significance. Despite such differences in address configuration, the textual descriptions exhibit a high degree of similarity. This observation underscores the rationale behind the main focus of this study on matching information points primarily through text.

4.2 Implementation

Our model incorporated both word and character vectors, utilizing pre-trained Word2Vec (Li et al., 2018) embeddings trained from the Chinese Wikipedia and the electrical vocabulary corpus extracted from the Sogou Input Method, it was then fine-tuned on our dataset for virtual terminal matching with a reduced learning rate of 0.0001 to prevent overfitting on the smaller dataset. The fine-tuning process lasted for 10 epochs. Each vector was set to a dimensionality of 300. To mitigate overfitting and enhance accuracy, Dropout probability was introduced during the experimentation. Following the input layer, sentence vectors are fed into a dual-layer bidirectional GRU with a hidden layer dimension of 128 for each GRU. The attention mechanism comprised eight units, and a rectified linear unit served as the activation function. To achieve optimal experimental results, early stopping was implemented as a training strategy.

In the experiments, the dimension of the word vectors used to initialize embedding vectors was 100, and we fixed the word embedding. The maximum length of the input sequence we chose was 15, and characters that were not in the dictionary were replaced with 0. The model was trained to minimize the cross-entropy of error loss through backpropagation and the Adam optimization algorithm was used with a 0.001 learning rate. The dropout rate was 0.5.

The experimental evaluation drew inspiration from concepts in machine learning, employing precision and recall. The comparative results can be categorized into true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The confusion matrix, as illustrated in Table 2, served as the basis for the performance evaluation using the accuracy and F1-score, as describe by Eqs 22–25:

$$recall\ rate = \frac{TP}{FN + FP} \tag{22}$$

$$precision\ rate = \frac{TP}{FP + TP} \tag{23}$$

$$accuracy = \frac{TN + FP}{FN + TN + FP + TP} \tag{24}$$

$$F1 - score = 2 * \frac{precision\ rate * recall\ rate}{precision\ rate + recall\ rate} \tag{25}$$

4.3 Baseline methods

In the experiment dataset, eight baseline methods were employed, including representation-based text classification

TABLE 2 Confusion matrix.

Reality	Matching result	
	Positive example	Negative example
True	TP	FN
False	FP	TN

models such as Bi-LSTM, CNN, TextCNN (Cao and Zhao, 2018); interactive-based text classification models such as ESIM (Chen et al., 2016), Siamese Bi-LSTM (Li et al., 2021), Attention-Bi-LSTM (Xie et al., 2019); and pre-trained based text classification models such as BERT.

In parameter settings, we strived to ensure consistency across various models as much as possible; in cases where consistency could not be guaranteed, efforts were made to maintain consistency with the original literature. Specifically, for CNN and TextCNN, a two-layer feedforward neural network was employed, with each layer having 128 hidden units. For Bi-LSTM, ESIM, Siamese Bi-LSTM, and Attention-Bi-LSTM, the hidden units for both the Bi-LSTM and feedforward neural network were set to 256. Additionally, consistent with both baseline methods and the proposed approach, Dropout (uniformly set to 0.5) was utilized to mitigate overfitting issues, while training was conducted using the Adam optimizer with a learning rate of 0.001. Regarding BERT, we first separated text pairs in the samples using the SEP token and then inputted them into the BERT-Base model. The vector corresponding to the [CLS] token at the head was extracted as the matching vector for the two sentences, which was then fed into a feedforward neural network to obtain the matching results for the two sentences. Due to convergence issues with a high learning rate, the learning rate for the BERT model was set to 0.0001.

4.4 Results and discussion

4.4.1 Comparative experiment

In order to substantiate the superiority of the proposed text matching model, which integrates multi-head attention and Siamese neural networks, experiments were conducted. The results of the comparisons are presented in Table 3.

The experimental results showcased in Table 3 highlight the effectiveness of our proposed text-matching model compared to traditional text classification approaches. Our proposed text matching model outperformed all baseline models, achieving an impressive accuracy of 97.52% and an F1-score of 97.89%. The integration of multi-head attention and Siamese neural networks enables our model to effectively capture semantic similarities between sentences. The substantial performance improvement of our model over traditional approaches underscores the importance of integrating multi-head attention and Siamese neural networks for text-matching tasks. The superior accuracy and F1-score achieved by our model signify its robustness and effectiveness in capturing semantic relationships between sentences.

Table 3 displayed the training times for nine experimental models on training and test dataset. Bi-LSTM, CNN,Attention-Bi-LSTM and TextCNN demonstrated a clear advantage in training time, while ESIM and Siamese Bi-LSTM required longer training times. Our model integrated Siamese network, Bi-GRU, and muti-head attention, resulting in a complex structure, hence its training time was only surpassed by the structurally complex BERT. In practical applications, since text matching model training generally occurs offline, the model's time complexity requirement is not high, with more emphasis placed on the accuracy of similarity judgment. Additionally, the training time of our model is essentially the same as that of the baseline Siamese Bi-LSTM model. This indicates that in the task of virtual terminal matching, the extraction of interaction features from text pairs based on interaction units has minimal impact on the model's time complexity.

4.4.2 Ablation experiment

In order to comprehensively understand the contribution of different aspects of our proposed model, we conducted an ablation study. We explored various granularities, pooling strategies (average, max), multi-head attention mechanisms, and the impact of incorporating address string validation on experimental results. The findings are summarized in Table 4.

The experimental results indicate that utilizing both character and word embeddings as input can capture more textual information. Employing both max-pooling and average-pooling facilitates effective interaction with semantic information in sentence pairs. The incorporation of attention mechanisms enables the model to capture diverse semantic relationships, thereby enhancing its performance. Additionally, integrating port address validation improves the model's accuracy and F1-score, ensuring its robustness in text matching tasks.

4.4.3 Parameter sensitivity experiment

In the experiment, the variation in the number of heads in the multi-head attention mechanism and the layers in the GRU has a certain impact on the model. Therefore, this study employs sensitivity analysis to investigate and analyze the parameters. Sensitivity analysis primarily involves analyzing the effect of changing a specified piece of information under the assumption of a certain state, designating it as the independent variable, and examining how this designated independent variable affects changes in other variables. In this experiment, we set the independent variables as the number of heads in the multi-head attention mechanism and the layers in the GRU, exploring their effects on the trend of virtual terminal matching results.

(1) Impact of the number of heads in multi-head attention

Multi-head attention enables the aggregation of information from multiple dimensions, facilitating a better understanding of semantic information from different spatial perspectives and preventing overfitting. Leveraging this characteristic, this study tests the number of heads in the multi-head attention mechanism on the validation set, sequentially setting the number of heads as [2, 4, 6, 8, 10] for experimentation. The most suitable number of heads is selected to configure parameters for the pseudo-anchor matching model, as shown in Figure 9 with the experimental results.

According to the experimental results in Figure 9, it is evident that on the validation set, when the number of heads in the multi-head attention mechanism reaches 8, the F1-score and accuracy

TABLE 3 Comparative experiment results.

Model	Accuracy (%)	F1-score (%)	Training time (s)
Bi-LSTM	82.64	83.15	2,310
CNN	83.69	84.27	983
TextCNN	93.77	93.89	1280
ESIM	94.13	94.49	3379
Siamese Bi-LSTM	95.56	95.67	3529
Attention-Bi-LSTM	94.02	94.06	2,581
BERT	96.17	96.28	62,788
Our model	97.52	97.89	3752

TABLE 4 Ablation experiment results.

Model	Accuracy (%)	F1-score (%)
Only word	93.78	93.89
Only character	94.16	94.53
Average pooling	95.23	95.43
Max pooling	95.31	95.39
Without multi-head attention	95.87	96.18
Without port address validation	94.41	94.93
Our model	97.52	97.89

attain their maximum values at 85.01. At this point, the model demonstrates optimal performance. As the number of heads gradually increases or decreases from 8, the F1-score and accuracy of the model progressively decreases. When a single attention head is applied to a sentence, although each word embedding contains embeddings from other words, it is predominantly influenced by the embedding of the word itself. However, utilizing multiple heads enables attention to be concentrated at different positions, aggregating multi-layered information. The advantage of the multi-head attention mechanism lies in its ability to balance the model, providing an additional space for parameter adjustment. This indicates that neither an excessive nor insufficient number of heads is optimal; rather, a balance is required. Both an excess or a deficiency in the number of heads will affect the virtual terminal matching results. Through testing, it is observed that in this model, the optimal performance in virtual terminal matching is achieved when the number of heads is set to 8.

(2) Impact of GRU layers

GRU strengthens the connection between vocabulary and context when processing information, enriching the semantic information of features and alleviating the problem of differences between sentences. The different layers of GRU affect the complexity of the model and have a certain impact on data fusion. In this study, GRU is utilized to process information. To investigate the influence of GRU layers on the model, experiments are set up, selecting GRU

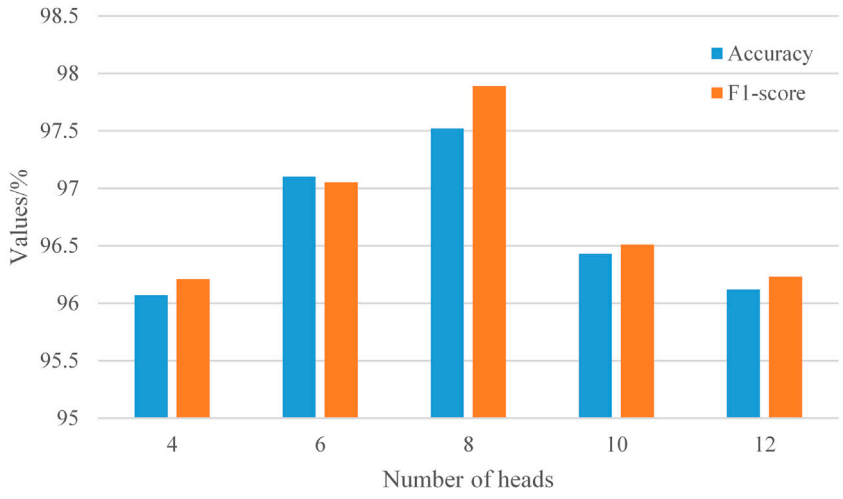


FIGURE 9
Accuracy and F1-score with different number of heads.

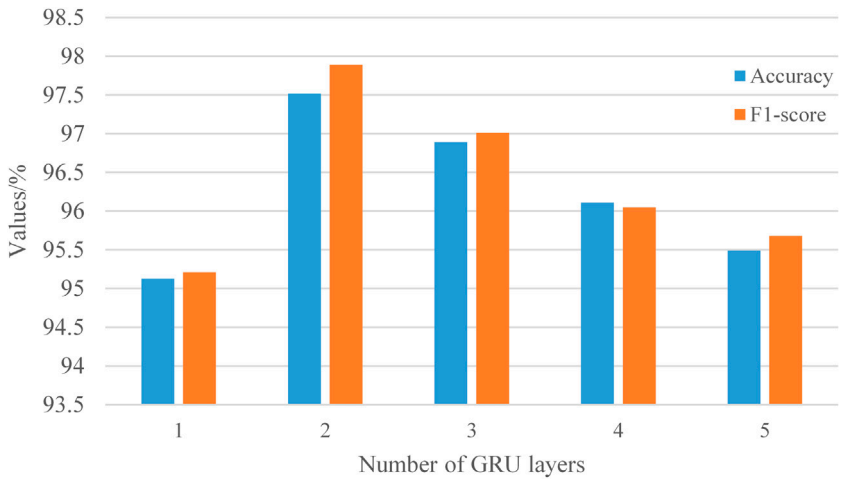
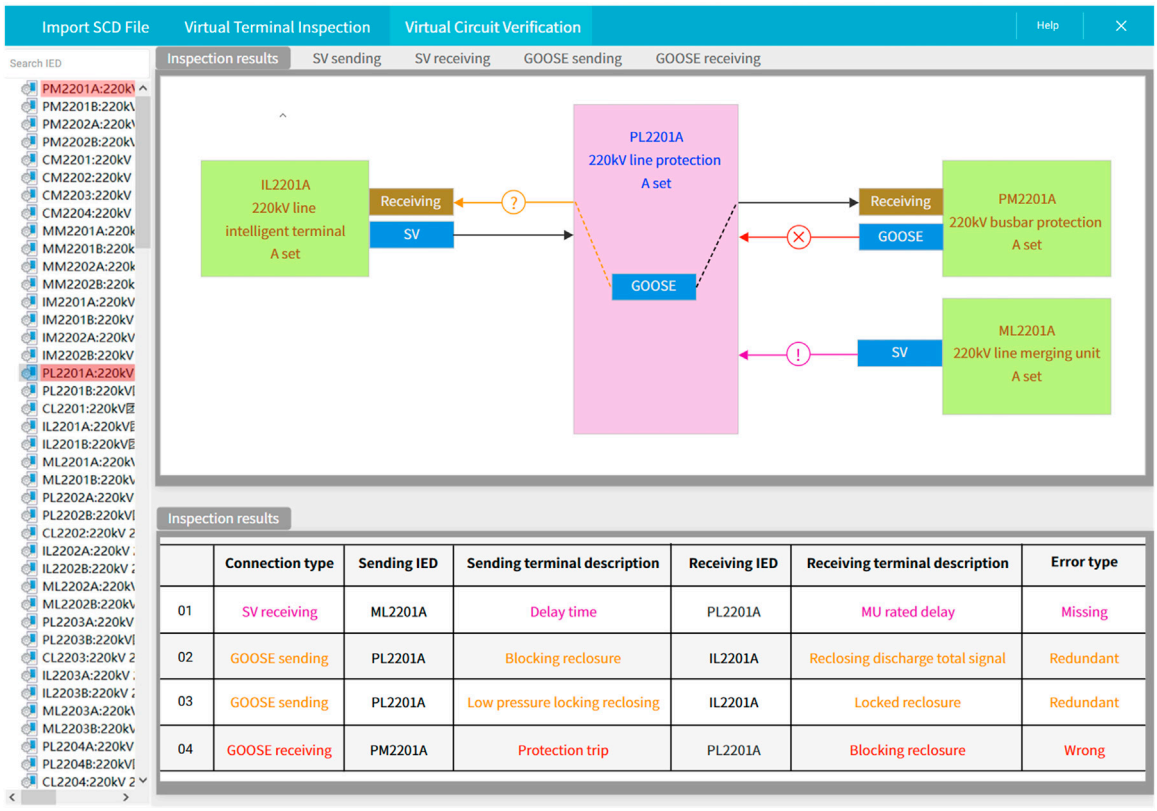


FIGURE 10
Accuracy and F1-score with different number of GRU layers.



	Connection type	Sending IED	Sending terminal description	Receiving IED	Receiving terminal description	Error type
01	SV receiving	ML2201A	Delay time	PL2201A	MU rated delay	Missing
02	GOOSE sending	PL2201A	Blocking reclosure	IL2201A	Reclosing discharge total signal	Redundant
03	GOOSE sending	PL2201A	Low pressure locking reclosing	IL2201A	Locked reclosure	Redundant
04	GOOSE receiving	PM2201A	Protection trip	PL2201A	Blocking reclosure	Wrong

FIGURE 11
Verification results of virtual circuit verification system.

TABLE 5 Comparison of verification results.

Verification method	Number of virtual circuits	Accuracy (%)	Time (min)
Manual verification	6000	92	160
Our model	6000	99	3

layers as [1, 2, 3, 4, 5], and testing and analyzing them on the validation set. The specific results are shown in Figure 10.

From Figure 10, it can be observed that when the number of GRU layers is set to 2, the model performs better compared to other values. This is primarily because when the number of GRU layers is 1, the model can only learn information in one direction, resulting in poor fitting to the dataset and inferior performance in extracting information compared to multi-layer GRU. However, when the number of layers increases to 2, GRU can learn information in two directions, enabling better learning of forward and backward contextual information. This facilitates capturing deeper relationships among hidden states, enriching the textual representation of vocabulary, and obtaining better text representation. However, due to the existence of GRU, both memory and time overheads increase. Therefore, as the number of GRU layers continues to increase, the time overhead increases with the increase in the number of neurons in GRU, and memory overhead also becomes significant. Moreover, the problem of vanishing gradients between layers becomes more apparent, leading to issues with data generalization and a higher likelihood of overfitting. Hence, this study sets the number of GRU layers to two to better learn information from various aspects and achieve better virtual terminal matching results.

4.4.4 Engineering application

The SCD editing tool modifies the SCD file in the sample to be verified. This process manually sets the error virtual circuits. A virtual circuit verification system based on the established knowledge graph and virtual circuit verification process, the verification results are shown in Figure 11. When a virtual circuit does not exist in the verification template based on the SCD file that passed the verification, the program automatically identifies the newly added virtual circuit and marks it as "!" as a reminder. When a virtual circuit is missing, the program automatically identifies the missing virtual circuit and marks it "?" as a reminder. The standard terminal library and virtual circuit verification template file based on the proposed method automatically verify the virtual circuit of the SCD file. The verification results were correct, showing the effectiveness of the proposed method.

In order to demonstrate the feasibility of system application, the efficiency and accuracy of intelligent verification and verification designed in the article were compared with manual verification by selecting the same number of intelligent substation configuration file verification tasks. The comparison results are shown in Table 5. It can be intuitively observed that compared with manual verification, the application of automatic verification technology for virtual circuit verification can help improve the efficiency and accuracy of intelligent substation virtual circuit verification.

5 Conclusion

In addressing the complexities associated with the verification of virtual circuits in intelligent substations, this study introduces a novel method that synergizes the strengths of knowledge graphs and deep learning. Through this fusion, we not only enhance the accuracy of virtual circuit verification but also set a new benchmark that surpasses the capabilities of traditional manual inspections and existing automated solutions. Our approach, characterized by its innovative integration of a Siamese neural network with a multi-head attention mechanism, demonstrates robust performance in the context of virtual terminal matching. Additionally, the inclusion of virtual terminal address string verification further enhances the accuracy of virtual circuit verification, presenting a new method for the verification of virtual loops in intelligent substations.

Future research will explore the method's adaptability to real-time configuration changes and potential integration with existing substation management systems, aiming to provide a more cohesive and efficient operational framework for intelligent substations and the broader power system. In addition, as intelligent substations continue to evolve, incorporating advancements such as IoT devices and advanced communication protocols, future research will focus on adapting our verification method to these new technologies. This includes exploring how to effectively process and integrate real-time data from a variety of sources to continuously update and refine the verification process.

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

HC: Conceptualization, Writing–original draft. YZ: Methodology, Writing–original draft. YG: Methodology, Writing–original draft. JS: Methodology, Software, Writing–original draft. CT: Supervision, Validation, Writing–review and editing. XR: Supervision, Validation, Writing–review and editing. HC: Data curation, 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 State Grid Jiangsu Electric Power Co., Ltd. Incubation Project (JF2023011).

Conflict of interest

Authors HC, YG, CT, and XR were employed by State Grid Jiangsu Electrical Power Company. Authors YZ and HC were employed by State Grid Nanjing Power Supply Company. Author JS was employed by State Grid Suzhou Power Supply Company.

The authors declare that this study received funding from State Grid Jiangsu Electric Power Co., Ltd. The funder had the following involvement in the study: during the manuscript writing phase, they provided suggestions on the organization of the article. After the manuscript was written, they provided constructive comments on the way the images were presented. They were also involved in the revision of the manuscript.

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, A., Charalampopoulos, P., Pissis, S. P., and Radoszewski, J. (2020). Dynamic and internal longest common substring. *Algorithmica* 82, 3707–3743. doi:10.1007/s00453-020-00744-0

Bieder, F., Sandkühler, R., and Cattin, P. C. (2021) *Comparison of methods generalizing max-and average-pooling*.

Cao, Y. K., and Zhao, T. (2018). "Text classification based on TextCNN for power grid user fault repairing information," in *5th international conference on systems and informatics*, 1182–1187. doi:10.1109/ICSAI.2018.8599486

Chen, Q., Zhu, X. D., Ling, Z. H., Wei, S., Jiang, H., and Inkpen, D. (2016) *Enhanced LSTM for natural language inference*.

Chen, X., Zhang, C., Liu, Q. K., Peng, Y., Zhou, D. M., and Zhen, J. M. (2021). Automatic configuration method of intelligent recorder based on deep semantic learning. *Power Syst. Prot. Control* 49 (02), 179–187. doi:10.19783/j.cnki.pspc.200367

Chen, X. J., Jia, S. B., and Xiang, Y. (2020). A review: knowledge reasoning over knowledge graph. *Expert Syst. Appl.* 141, 112948. doi:10.1016/j.eswa.2019.112948

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014) *Learning phrase representations using RNN encoder-decoder for statistical machine translation*.

Cui, Y. G., Zhang, F., Liu, H. Y., Ji, J., Liu, X. G., and Zhao, J. (2018). "Research on secondary circuit identification technology and condition-based maintenance mode of intelligent substation," in *2018 international conference on power system technology*, 3716–3722. doi:10.1109/POWERCON.2018.8602073

Fan, W. D., Feng, X. W., Dong, J. X., Hu, Y., Wang, W. Q., and Gao, X. (2020). Automatic matching method of a virtual terminal in intelligent substation based on semantic similarity of historical data. *Power Syst. Prot. Control* 48, 179–186. doi:10.19783/j.cnki.pspc.191272

Guan, N. N., Song, D. D., and Liao, L. J. (2019). Knowledge graph embedding with concepts. *Knowledge-Based Syst.* 164, 38–44. doi:10.1016/j.knosys.2018.10.008

Guo, Q., Huang, J., Xiong, N., and Wang, P. (2019). MS-pointer network: abstractive text summary based on multi-head self-attention. *IEEE Access* 7, 138603–138613. doi:10.1109/ACCESS.2019.2941964

Hao, X. G., Zhao, Y. H., Yin, X. G., Luo, P., Yang, J. C., and Mao, Y. R. (2020). Intelligent substation virtual circuit check based on the intermediate model file. *J. Electr. Power Sci. Technol.* 35, 132–137. doi:10.19781/j.issn.16739140.2020.05.018

Huang, Q., Jing, S., Li, J., Li, J., Cai, D. S., Wu, J., et al. (2017). Smart substation: state of the art and future development. *IEEE Trans. Power Deliv.* 32, 1098–1105. doi:10.1109/TPWRD.2016.2598572

Li, C., and Wang, B. (2023). A knowledge graph method towards power system fault diagnosis and classification. *Electronics* 12, 4808. doi:10.3390/electronics12234808

Li, S., Zhao, Z., Hu, R. R., Li, W. S., Liu, T., and Du, X. Y. (2018) *Analogical reasoning on Chinese morphological and semantic relations*.

Li, Z. G., Chen, H., and Chen, H. Y. (2021). Biomedical text similarity evaluation using attention mechanism and Siamese neural network. *IEEE Access* 9, 105002–105011. doi:10.1109/ACCESS.2021.3099021

Liang, G., On, B. W., Jeong, D., Kim, H. C., and Choi, G. S. (2018). Automated essay scoring: a siamese bidirectional LSTM neural network architecture. *Symmetry* 10 (12), 682. doi:10.3390/sym10120682

Liu, R., Fu, R., Xu, K., Shi, X. Z., and Ren, X. N. (2023). A review of knowledge graph-based reasoning technology in the operation of power systems. *Appl. Sci.* 13, 4357. doi:10.3390/app13074357

Oliveira, B. A. S., De Faria Neto, A. P., Fernandino, R. M. A., Carvalho, R. F., Fernandes, A. L., and Guimarães, F. G. (2021). Automated monitoring of construction sites of electric power substations using deep learning. *IEEE Access* 9, 19195–19207. doi:10.1109/ACCESS.2021.3054468

Ren, J. B., Li, T. C., Gen, S. B., Liu, Q. Q., He, Y. K., Wang, Z. H., et al. (2020). An automatic mapping method of intelligent recorder configuration datasets based on Chinese semantic deep learning. *IEEE Access* 8, 168186–168195. doi:10.1109/ACCESS.2020.3024060

Reyad, M., Sarhan, A. M., and Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. *Neural Comput. Appl.* 35 (23), 17095–17112. doi:10.1007/s00521-023-08568-z

Selim Ustun, T., and Suhail Hussain, M. (2020). An improved security scheme for IEC 61850 mms messages in intelligent substation communication networks. *J. Mod. Power Syst. Clean* 8, 591–595. doi:10.35833/MPCE.2019.000104

Song, Q. P., Sheng, W. X., Kou, L. F., Zhao, D. B., Wu, Z. P., Fang, H. F., et al. (2016). Smart substation integration technology and its application in distribution power grid. *CSEE J. Power Energy* 2, 31–36. doi:10.17775/CSEEJPES.2016.00046

Tian, J. P., Song, H., Sheng, G. H., and Jiang, X. C. (2022). An event knowledge graph system for the operation and maintenance of power equipment. *IET Generation, Transm. Distribution* 16, 4291–4303. doi:10.1049/gtd2.12598

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. neural Inf. Process. Syst.* 30, 5999–6009. doi:10.48550/arXiv.1706.03762

Wang, Q., Mao, Z. D., Wang, B., and Guo, L. (2017). Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans. Knowl. Data En.* 29 (12), 2724–2743. doi:10.1109/TKDE.2017.2754499

Wang, W. Q., Hu, Y., Zhao, N., and Wu, C. Y. (2018). Automatic connection method of virtual terminators based on optimization model of distance weight vectors. *Power Syst. Technol.* 42 (01), 346–352. doi:10.13335/j.1000-3673.pst.2017.1746

Xie, J., Chen, B., Gu, X. L., Liang, F. M., and Xu, X. Y. (2019). Self-attention-based BiLSTM model for short text fine-grained sentiment classification. *IEEE Access* 7, 180558–180570. doi:10.1109/ACCESS.2019.2957510

Zhang, X. L., Liu, H. H., Li, J. Q., Ai, J. Y., Tang, Y., and Wang, J. Y. (2015). Automatic test platform in smart substation for relay protection. *Automation Electr. Power Syst.* 39 (18), 91–96. doi:10.7500/AEPS20141103006

Zhao, Z. C., Lu, D. H., Liu, S. J., Chen, S., and Ding, J. W. (2023). Research on automatic correlation technology of virtual circuit in secondary system of intelligent substation. *J. Phys. Conf. Ser.* 2479, 012068. doi:10.1088/1742-6596/2479/1/012068

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 14 June 2024
DOI 10.3389/fenrg.2024.1409647

Check for updates

OPEN ACCESS

EDITED BY
Jinran Wu,
Australian Catholic University, Australia

REVIEWED BY
Linfei Yin,
Guangxi University, China
Du Zongjuan,
Xi'an Jiaotong University, China

*CORRESPONDENCE
Yuze Ji,
jcjiyuze@163.com

RECEIVED 30 March 2024
ACCEPTED 02 May 2024
PUBLISHED 14 June 2024

CITATION
Chen W, Wang X, Ji Y, Zhang Y, Zhu J and Ma W (2024), A physical virtual multi-graph convolutional coordinated prediction method for spatio-temporal electricity loads integrating multi-dimensional information. *Front. Energy Res.* 12:1409647. doi: 10.3389/fenrg.2024.1409647

COPYRIGHT
© 2024 Chen, Wang, Ji, Zhang, Zhu and Ma. 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 physical virtual multi-graph convolutional coordinated prediction method for spatio-temporal electricity loads integrating multi-dimensional information

Wengang Chen, Xinrui Wang, Yuze Ji*, Yujuan Zhang, Jianfei Zhu and Weitian Ma

Jincheng Power Supply Company, State Grid Shanxi Electric Power Company Limited, Jincheng, China

Traditional load prediction methods are unable to effectively predict the loads according to the spatial topology of each electricity consumer in neighboring areas and the load dependency correlations. In order to further improve the load prediction accuracy of each consumer in the region, this paper proposes a short-term prediction method of electric load based on multi-graph convolutional network. First, the input data are selected with maximum information coefficient method by integrating multi-dimensional information such as load, weather, electricity price and date in the areas. Then, a gated convolutional network is used as a temporal convolutional layer to capture the temporal features of the loads. Moreover, a physical-virtual multi-graph convolutional network is constructed based on the spatial location of each consumer as well as load dependencies to capture the different evolutionary correlations of each spatial load. Comparative studies have validated the effectiveness of the proposed model in improving the prediction accuracy of power loads for each consumer.

KEYWORDS

graph convolutional network, short-term load, multidimensional information, spatiotemporal prediction, maximum information coefficient

1 Introduction

The global electricity demand is experiencing rapid growth, and the structure of urban distribution networks is becoming increasingly complex, which elevates the challenges associated with power grid scheduling and control (Hou et al., 2021). The ongoing expansion of hybrid renewable power systems has led to the integration of a substantial number of variable renewable energy sources, such as wind and solar, transforming the grid into an active distribution network. This transformation has concomitantly increased the volatility and uncertainty inherent to power systems (Cleary et al., 2015). Accurate load prediction is of paramount importance for enhancing the safety, stability, and efficient operation of the power grid (Celebi and Fuller, 2012). Furthermore, as power systems undergo reform, electricity sales companies and virtual power plants participating in the electricity market

must accurately predict the electricity consumption of individual consumers (Aparicio et al., 2012).

The existing research on power load prediction can be broadly categorized into two primary methodological approaches: statistical models and machine learning techniques. The statistical modeling approach offers simplicity and expedient prediction capabilities. Prominent statistical methods include linear regression and exponential smoothing (Shi W. et al., 2023). Saber and Alam (2017) leveraged the autoregressive integrated moving average model to analyze the correlation between load demand and influential factors, and established a non-stationary stochastic prediction framework. However, such statistical techniques generally suffer from limitations in prediction accuracy and robustness. In contrast, machine learning methods possess adaptive and self-learning capabilities that have demonstrated improvements in load prediction precision. These advanced analytical techniques include support vector machines, extreme learning machines, long short-term memory (LSTM) networks, and convolutional neural networks (Li et al., 2020; Samadianfard et al., 2020; Zhang J. et al., 2021; Tang et al., 2021; Roy and Yeafi, 2022; Sun et al., 2023; Deng et al., 2024). Specifically, Li et al. (2020) proposed a power load decomposition and reconstruction prediction approach based on support vector machines. Furthermore, Roy and Yeafi (2022) and Sun et al. (2023) leveraged machine learning theory to establish residual self-attention encoding-decoding networks for electricity consumption and wind power prediction, effectively capturing the coupling relationships within the data. Additionally, Tang et al. (2021), Zhang J. et al. (2021) and Samadianfard et al. (2020) employed echo state networks, LSTM, and multi-layer perceptrons to predict wind direction, speed, and power generation. While the aforementioned methods utilize multi-dimensional information, such as load data and weather factors, to model the temporal correlations in load patterns, they have largely overlooked the potential spatial correlations in electricity consumption among multiple consumers. Neighboring consumers are affected by factors such as weather, electricity prices, and holidays, exhibiting similar electricity consumption behaviors and load profiles (Lin et al., 2021). Fully capturing and leveraging the spatial correlation information among these neighboring consumers has the potential to further improve the accuracy of load prediction. However, the non-Euclidean, interconnected graph structure of the consumer data limits the direct applicability of conventional neural network architectures, and thus necessitates specialized modeling approaches capable of learning from the complex spatial correlation of neighboring consumers.

Graph neural networks have attracted widespread attention because they can learn implicit representations of node data on graph structures and process non-Euclidean spatial data. Currently, graph neural networks have been successfully applied in fields such as transportation and load prediction. Yan et al. (2021) proposed a multi-time scale traffic prediction method based on graph convolutional networks, which treats each road sensor as a node to construct a spatio-temporal module and capture spatio-temporal correlations. Liao et al. (2023) established a three-dimensional Gaussian wake function that represents the relevant information of each wind turbine and used graph neural networks combined with attention mechanisms to predict the output power of non-uniform wind farms, reducing prediction errors. Shi P. et al. (2023) proposed

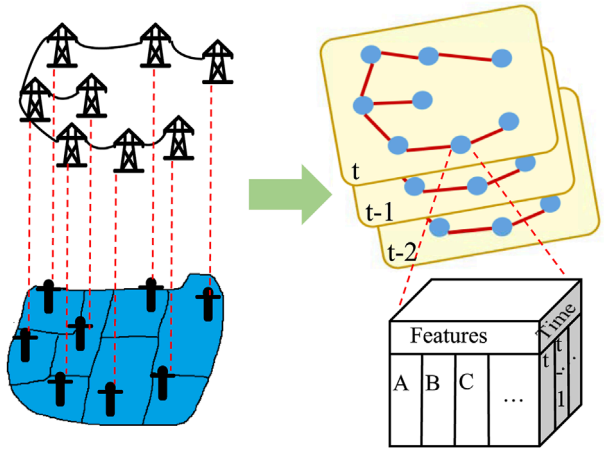


FIGURE 1
Spatial and temporal load structure diagram.

a multi-user short-term power load spatio-temporal prediction method using multi-head attention and adaptive graph theory and compared it with various methods. Zhang L. et al. (2021) used K-means clustering to divide user groups, capture the intrinsic spatio-temporal correlation information of the data using local spatio-temporal graphs, and finally aggregate the calculation results of each part to predict the future spatio-temporal power demand sequence. Fahim et al. (2024) took a load of each charging station as a node, used an adaptive adjacency matrix to reflect the spatial relationship between stations, and proposed a multi-station charging demand prediction method for electric vehicle charging stations based on graph networks. Existing literature has shown that graph neural networks can explore potential relationships between loads and improve prediction accuracy. However, the above literature only considers the fixed spatial connection relationships of consumers and relies on a single graph representation, failing to reflect the various spatial correlations between electricity loads in the neighbors.

Multimodality neural networks have improved prediction accuracy, which has attracted the attention of researchers. Zheng et al. (2023) used virtual dynamic graph and physical road graph to extract heterogeneous, variable, and inherent spatial patterns of the road network. Liu et al. (2020) presented a physical-virtual collaboration graph neural network for passenger flow prediction. The network is a general model that can be directly applied to online pedestrian flow prediction. Xiu et al. (2024) adopted parallel convolutional networks and combines relational data within the metro network to predict ridership. In addition, the train timetable as feature input to the network, improving prediction accuracy. However, there is limited application of literature in load prediction. How to design a proper prediction network based on the characteristics of electricity demands is an important issue.

Multi-graph convolutional networks have been applied in literature for load prediction. Wei et al. (2023) presented a novel multi-graph neural networks for short-term electricity demand prediction, which is embedded with the directed static graph and directed dynamic graph. The results show that the network has

TABLE 1 Influencing factors of electric load.

Type	Factor	Meaning	Label
Historical load	Load	Historical electricity load	A
Time	Time	Time	B
	Date	Weekday, weekend, holiday	C
	Temperature	Temperature at the time of prediction	D
Meteorological condition	Humidity	Humidity at the time of prediction	E
	Wind speed	Wind speed at the time of prediction	F
	Rain	Rain at the time of prediction	G
Price	Electricity price	Peek, flat, valley	H

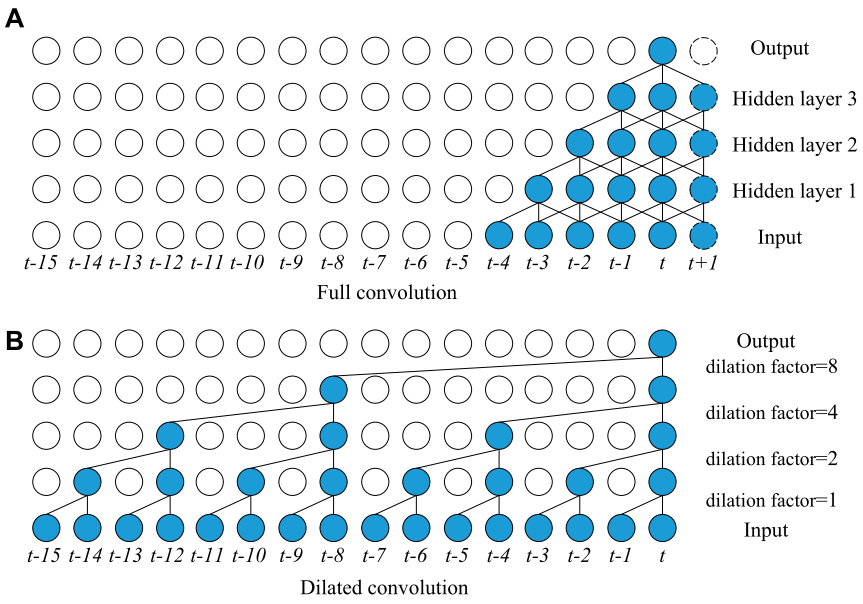


FIGURE 2
Convolutional network: **(A)**Full convolution, **(B)**Dilated convolution.

a strong ability to capture periodic features. Yanmei et al. (2024) adopted dynamic load knowledge graph to extract the correlation between internal at-tributes and external influencing factors of various loads. Moreover, the attention mechanism enhances the learning ability of load feature representation. To capture complex non-linear correlations of loads, Wang et al. (2023) proposed spatial and temporal graph neural network for residential load prediction. The multiple dependence graphs consists of synchronization graph and causality graph, which can model linear and non-linear dependence. However, the exist research has not fully captured the in-dependence with multidimensional data, and electricity demand is associated with various complex and unknown factors. Therefore, predefined graphs cannot fully reflect load correlations. In addition, the coupling of spatio-temporal multidimensional information and the large amount of data make effective utilization to improve model performance another key issue.

This article proposes a multi-graph convolutional spatio-temporal collaborative prediction method for power load integrating multi-dimensional information. By constructing a multi-graph network, the spatial information of each consumer's load is fully captured to improve prediction accuracy. First, based on historical load, weather, and electricity prices, the maximum information coefficient (MIC) is used to analyze the correlation of load sequence and to construct input data that integrates multi-dimensional information. Then, the network adopts a dilated convolution and gating mechanism to parallelly capture the practical information of temporal loads. Moreover, based on the actual location connection between consumers and the similarity of electricity loads, a physical, virtual multi-graph convolutional model is established to capture various interrelationships between loads in space. T the performance of the proposed model is tested on real

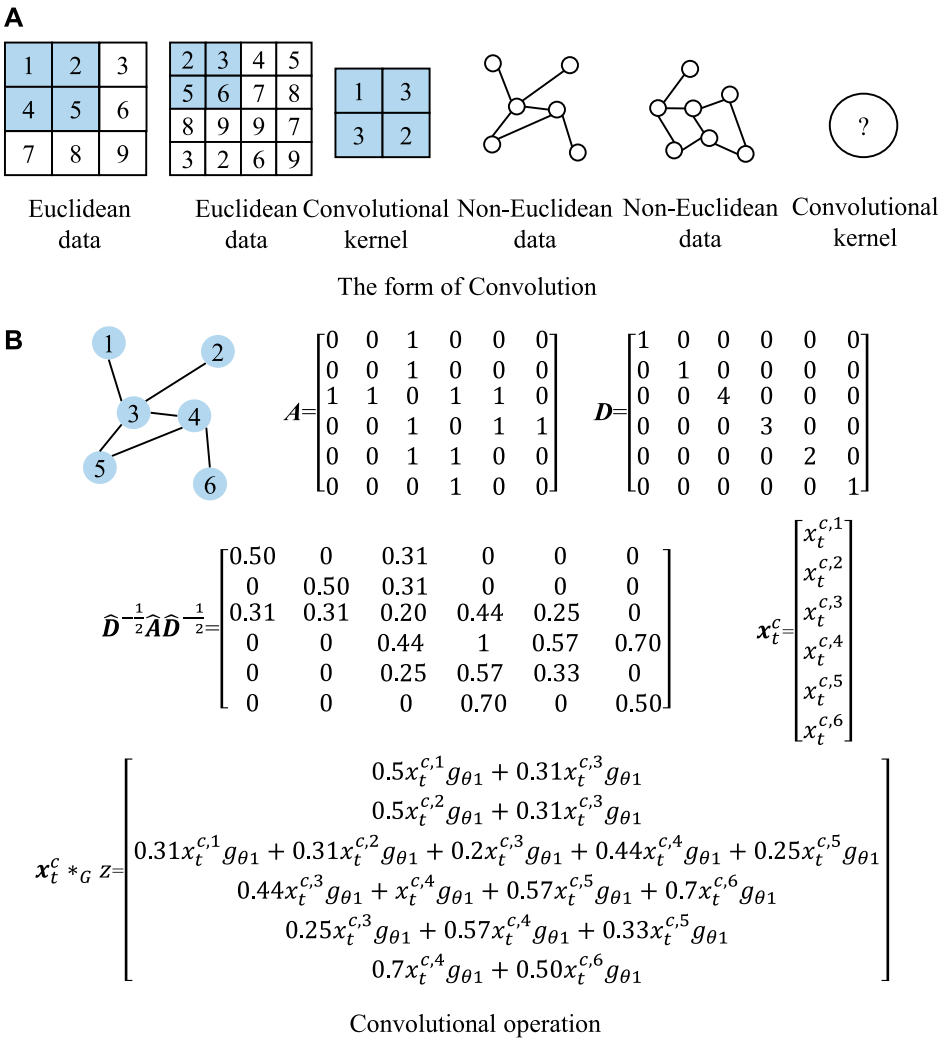


FIGURE 3
Graph convolution operation: **(A)**The form of convolution, **(B)**Convolutional operation.

electricity datasets and compared with other baseline models to verify its effectiveness.

The innovation of this paper is as the following:

- The gated casual convolution is adopted to accelerate the temporal convolution, which can capture correlation of time series information.
- We proposed a physical virtual multi-graph convolutional network to fully capture electricity load evolution patterns. The physical graph contains connection and distance data, which is based on realistic grid topology. The virtual graphs are built based on human domain knowledge.

Moreover, the main contribution of this paper is to use MIC to obtain the correlation of nonlinear influencing factors, which reduces input data redundancy. Specifically, this method filters out irrelevant spatio-temporal data and selects high MIC values as input, reducing the interference of input on the prediction results.

2 Spatio-temporal network of electricity loads

2.1 Spatial and temporal load structure

Stable electricity promotes the development of social production, and electricity is transmitted through power grid lines. The power consumption from different spatial locations is ultimately integrated into the power load of different grid nodes.As shown in Figure 1, different nodes in the grid topology correspond to electricity demand generated in different actual geographic areas. The load in an area corresponding to a grid node is regarded as the information of the nodes in the graph G, and the connection between grid nodes is regarded as the edges between the nodes in the graph G. We use the graph $G(V,E,A)$ to describe the spatial load information, where V is a node, E is an edge, and A is an adjacency matrix, representing the connection between nodes. Each node in the graph G generates data with a total number of features F in a time interval. As shown in Figure 1, each time slice is a spatial graph that records the feature information of all nodes in the time interval.

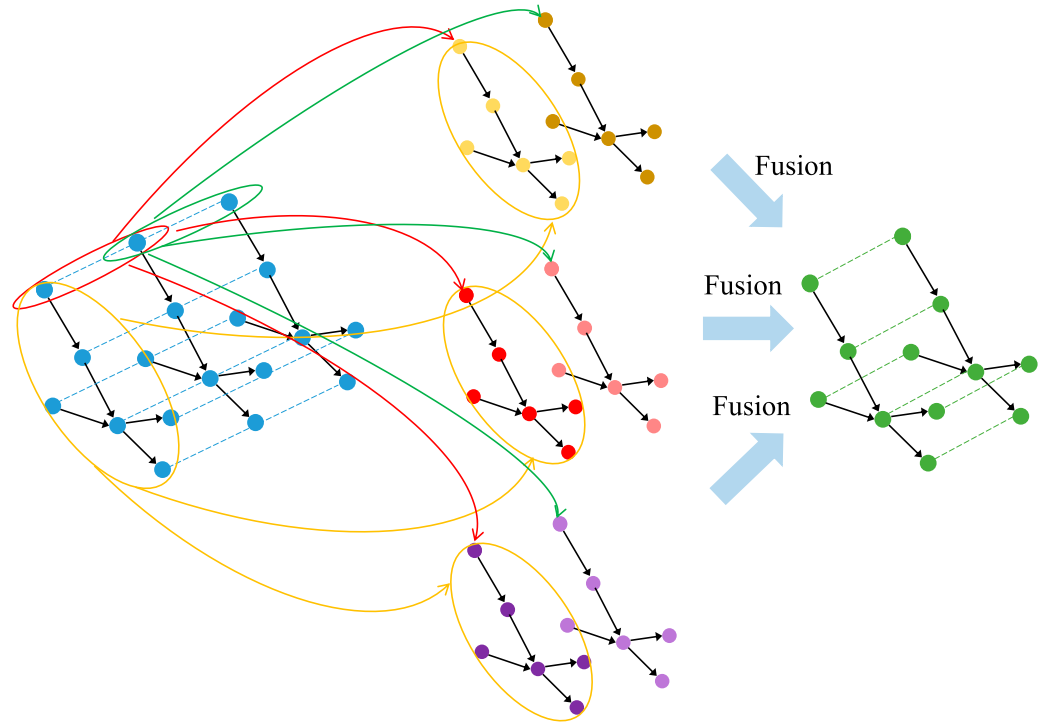


FIGURE 4
Spatio-temporal convolution.

2.2 Definition of parameters and sets for electricity load prediction

Let the electricity load generated by each node in the graph in a future time period be the forecast data. Let $x_t^{f,i}$ be the features f generated by node i in the time interval t, then $\boldsymbol{X}^c = (\boldsymbol{x}_1^f, \ldots, \boldsymbol{x}_\tau^f)$ be the features f generated by all nodes in time period τ, and $\boldsymbol{\chi} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_\tau)$ be all features generated by all nodes in time period τ. It is assumed that the load $y_t^i = x_t^{f,i}$ generated by a node i at a certain time t, that is, all the characteristics $\boldsymbol{\chi}$ including the load generated by all nodes within a certain time period are known. Then, the electricity load $\widehat{\boldsymbol{Y}} = (\hat{\boldsymbol{y}}^1, \ldots, \hat{\boldsymbol{y}}^N)$ for a certain time period in the future is predicted where $\hat{y}^i$ is the electricity load during a certain time period of the node.

2.3 Maximum information coefficient

The main factors influencing load prediction are historical loads, weather conditions, time, and electricity prices (Quilumba et al., 2014; Sun et al., 2022). Table 1 summarizes the load influencing factors. Although applying influencing factors directly to neural networks as input data can also predict loads, excessive data increases computational complexity and speed. Using proper methods to select input data can improve prediction accuracy and accelerate computational speed. Therefore, this paper applies the maximum information coefficient theory for feature extraction.

The maximum information coefficient was proposed by Reshef based on mutual information theory (Reshef et al., 2011). MIC can analyze the linear and nonlinear correlation between two variables and screen parameters that affect load. The mutual information between sequences X_a and Y_a can be expressed as Eq. 1.

$$I_m(X_a, Y_a) = \sum_{x_a \in X_a} \sum_{y_a \in Y_a} p(x_a, y_a) \log_2 \frac{p(x_a, y_a)}{p(x_a) p(y_a)} \tag{1}$$

where $I_m(X_a, Y_a)$ represents mutual information, and $p(\cdot)$ is the probability density function, $x_a \in X_a$ and $y_a \in Y_a$.

Let $D_a = \{(x_{a,i}, y_{a,i}), i = 1, \ldots, n\}$ be the set of binary data, and divide the value domains of X_a and Y_a into segments p_a and q_a in grid G_a. Define the maximum mutual information of D_a in grid G_a to be I_{mi} that is calculated using the Eq. 2.

$$I_{mi}(D_a, p_a, q_a) = \max I_m(D_a \mid G_a) \tag{2}$$

where $D_a \mid G_a$ represents the data D_a divided by grid G_a.

Therefore, the maximum information coefficient is formulated as Eq. 3.

$$I_{mic}(X_a, Y_a) = \max_{p_a q_a < B(n)} \frac{I_{mi}(D_a, p_a, q_a)}{\log_2(\min(p_a, q_a))} \tag{3}$$

where B_n is the limit on the number of grid divisions, generally $B_n = n^{0.6}$ (Reshef et al., 2011).

3 Spatio-temporal multi-graph prediction network

The spatio-temporal power prediction model mainly comprises a data embedding layer, a spatio-temporal prediction layer, and an output layer. The spatio-temporal prediction layer contains a

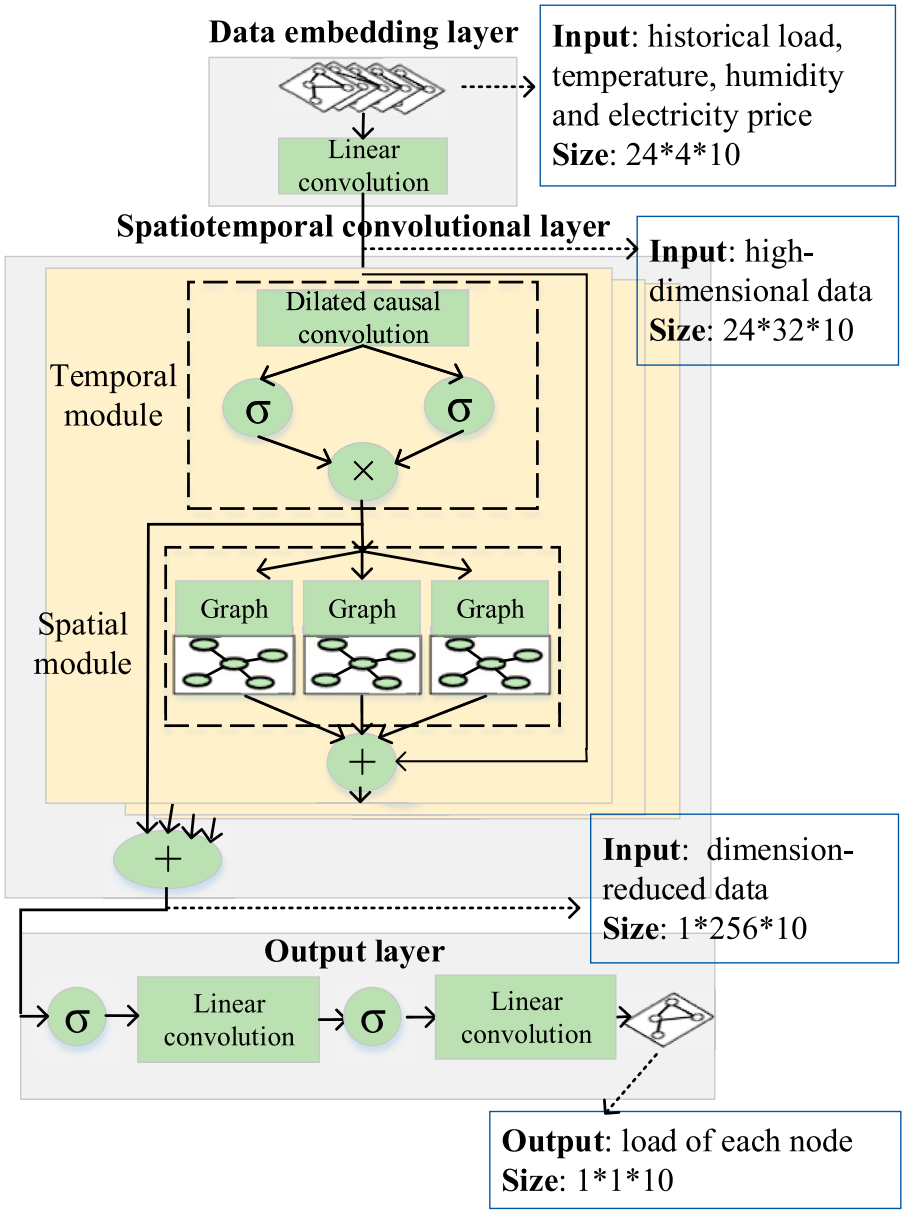


FIGURE 5
Spatio-temporal multi-graph prediction network structure.

temporal convolution module and a spatial multi-graph convolution module to capture the spatio-temporal correlation features of the data and the spatio-temporal dependencies of the data.

3.1 Data embedding layer

The data embedding layer consists of a convolutional network that transforms the input feature volume into high-dimensional data suitable for the spatio-temporal prediction layer. A standard convolutional network consists of three parts: convolutional, pooling, and fully connected layers, where the convolution is defined by Eq. 4

$$y_c = f(x_{in}^{*}\omega_c + b_c) \tag{4}$$

where x_{in} is the input to the convolutional layer; ω_c is the convolutional kernel, i.e., the weight parameter; b_c is the bias value; $f(\cdot)$ is the convolution operation; is the activation function; y_c is the output value. In this paper, linear convolution is used to linearly transform the input data into high dimensional data by convolution operation, i.e., no activation function is used.

3.2 Spatio-temporal prediction layer

Mining the dependencies of loads in the time dimension can help improve prediction accuracy, and choosing an appropriate network structure is crucial. Recurrent neural networks have the structure of loops that accept data from themselves and other neurons and are particularly suitable for processing time-series

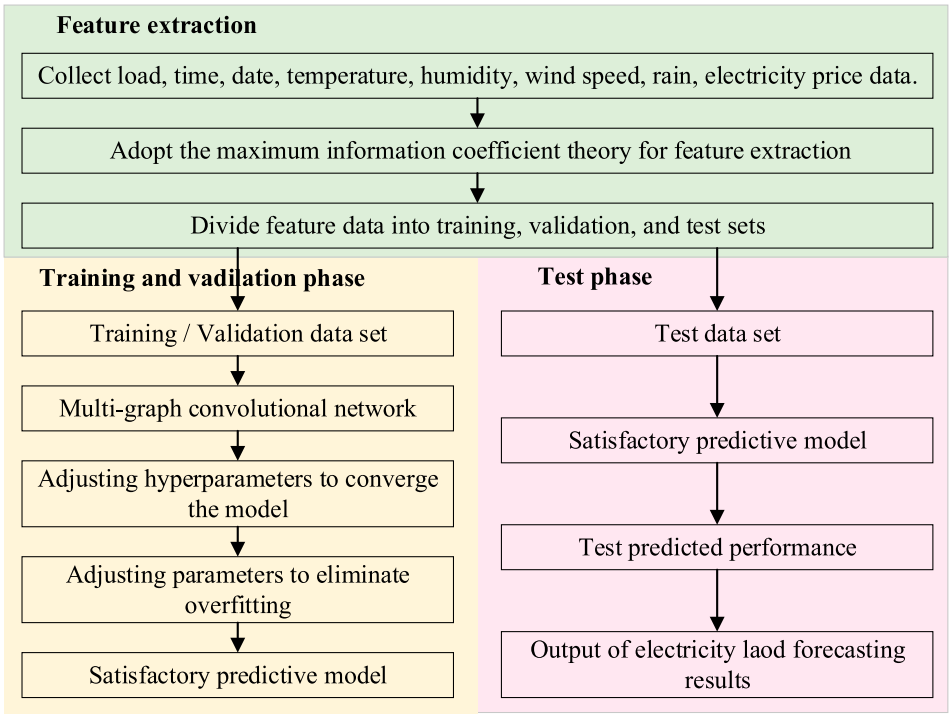


FIGURE 6
The overall flowchart.

TABLE 2 Parameters of the model.

Parameters	Values	Parameters	Values
Chebyshev polynomial order	3	Batch size	32
Iterations	200	Learning rate	0.0001
Dropout	0.3	Dilation factor	1, 2, 4, 8
Convolution kernel	2	Optimizer	Adam

data. However, deeper networks take a long time to compute results, and they are prone to gradient explosion and vanishing problems. In this paper, we choose convolutional neural networks with strong robustness and faster computation to capture the mutual characteristics of data in time.

3.2.1 Causal convolutional networks

Causal convolution is a special convolutional neural network that utilizes only past data in its computation. Its expansion factor can be controlled to quickly increase the receptive field, thus capturing load data for a longer period. As shown in Figure 2, causal convolution does not rely on data from future moments for computation compared to ordinary convolutional networks. In addition, stacking more layers of the null convolution can result in an exponential increase in the receptive field, covering more input data and speeding up the computation. The causal convolution can be expressed by Eq. 5 (Wu et al., 2019).

$$\boldsymbol{m}(t)^{*}\boldsymbol{x}_{\mathrm{d}}=\sum_{s=0}^{K-1}\boldsymbol{m}(s)\boldsymbol{x}_{\mathrm{d}}\left(t-q_{\mathrm{d}}s\right) \tag{5}$$

where $\boldsymbol{m}$ is the convolution kernel of the null convolution; s is the serial number of the convolution kernel; K is the size of the convolution kernel; $x_{\mathrm{d}} \in \mathbf{R}^{T}$ is the input sequence; t is the moment; q_{d} is the dilation factor, i.e., the interval between two factors.

3.2.2 Gated mechanisms

Gating methods can selectively control the rate of data accumulation to avoid memory saturation. Combining the gating mechanism with a casual convolutional network can capture the complex relationship between loads in the temporal dimension, which has a significant advantage in processing sequential data. The output of the gating operation can be expressed as Eq. 6.

$$\boldsymbol{h}_{\mathrm{g}}=g_{\mathrm{h}}\left(\boldsymbol{\omega}_{\mathrm{h}}^{*}\boldsymbol{x}_{\mathrm{g}}\right)\odot g_{\mathrm{s}}\left(\boldsymbol{\omega}_{\mathrm{s}}^{*}\boldsymbol{x}_{\mathrm{g}}\right) \tag{6}$$

where $\boldsymbol{x}_{\mathrm{g}}$ is the input data; $\boldsymbol{\omega}_{\mathrm{h}}$ and $\boldsymbol{\omega}_{\mathrm{s}}$ are the learnable model parameters; $g_{\mathrm{h}}(\cdot)$ is the hyperbolic tangent function; $g_{\mathrm{s}}(\cdot)$ is the Sigmoid function; $\odot$ is the operator for multiplying elements; $\boldsymbol{h}_{\mathrm{g}}$ is the output value. The temporal convolution module mined the features and correlations between power loads of the time series using gated null convolution network and fed the processed data into the spatial multi-graph convolution module.

3.3 Spatial multi-graph convolution module

In order to exploit the dependencies between electric loads in space, this section proposes a representation and calculation method of spatial loads based on spectral graph theory. Then, a physical-virtual multi-graph convolutional network based on the spatial location of loads and load similarity is built to represent the different dependencies of loads in the spatial dimension.

3.3.1 Spectral convolution

The graph load data is non-Euclidean space data, and each load node has a different connection relationship with other nodes. Moreover, the convolutional network is based on the translation invariant operation of the data, which cannot be directly applied to the non-Euclidean space. Bruna et al. (2014) defined graph convolution operation in spectral space based on graph theory and expressed the graph structure as a mathematical form. As a result, the non-Euclidean space data is transformed into Euclidean data for convolution operation.

In spectral theory, graph information can be represented by a Laplace matrix $\boldsymbol{L}$. The equation is $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{A}$ and the standard form is $\boldsymbol{L}=\boldsymbol{I}_{\mathbb{N}}-\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{\frac{1}{2}}$ where $\boldsymbol{A}$ is the adjacency matrix, $\boldsymbol{D}$ is the degree matrix, and $\boldsymbol{I}_{\mathrm{N}}$ is the identity matrix. Let $\alpha_{i,j}$ and $d_{i,j}$ be the elements of $\boldsymbol{A}$ and $\boldsymbol{D}$, and i and j be the number of rows and columns, then $d_{i,j}=\sum_j a_{i,j}$. Decompose L into eigenvalues, $\boldsymbol{L}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, where $\boldsymbol{U}$ is the eigenvector and $\boldsymbol{\Lambda}$ is the diagonal array of eigenvalues. The graph convolution operation relies on the Fourier transform, defining the Fourier transform of the information on the graph x as $\widehat{x}=\boldsymbol{U}^T x$ and $\widehat{x}$ as the value of x in the spectral domain. Accordingly, the Fourien inverse transform of x is $x=\boldsymbol{U}\widehat{x}$. By the convolution theorem, the Fourier transform of the convolution of signals is equal to the product of their individual Fourier transforms (Shuman et al., 2013). Therefore, it is possible to multiply the Fourier transform of the information on the graph by the information in the spectral domain and then invert the transformation to obtain the convolution result on the graph, as shown in Eq. 7.

$$x *_G z=\boldsymbol{U}\left(\left(\boldsymbol{U}^T x\right)\odot\left(\boldsymbol{U}^T y\right)\right)=\boldsymbol{U}\boldsymbol{g}_\theta\boldsymbol{U}^T x \tag{7}$$

where x and z are the signals on the graph; $*_G$ is the graph convolution; $\boldsymbol{g}_\theta$ is the convolution kernel, and $\boldsymbol{g}_\theta=\boldsymbol{U}^T z$

The graph convolution operation can be realized based on Eq. 7. However, calculating the Laplace matrix is cumbersome when the graph size is large. Therefore, the Chebyshev graph convolution approximation is used to solve the convolution kernel to simplify the operation:

$$x *_G z=\boldsymbol{U}\boldsymbol{g}_\theta\boldsymbol{U}^T x\approx\sum_m^{M-1}\theta_m T_m(\tilde{\boldsymbol{L}})x \tag{8}$$

where θ_m is the Chebyshev polynomial coefficients; $T_m(\tilde{\boldsymbol{L}})$ is the Chebyshev polynomial, $T_m(\tilde{\boldsymbol{L}})=2\tilde{\boldsymbol{L}}T_{m-1}(\tilde{\boldsymbol{L}})-T_{m-2}(\tilde{\boldsymbol{L}})$ and $T_0(\tilde{\boldsymbol{L}})=1$, $T_1(\tilde{\boldsymbol{L}})=\tilde{\boldsymbol{L}}$; $\tilde{\boldsymbol{L}}=2\boldsymbol{L}/\lambda_{\max}-\boldsymbol{I}_N$; and m is the order of Chebyshev polynomial.

The information in the graph is updated by the order information of itself and its neighboring nodes $M-1$, and the depth of the transmitted information can be adjusted by controlling the maximum order M. In the actual calculation, the value of $\boldsymbol{L}=\boldsymbol{I}_{\mathrm{N}}-\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}$ is dispersed, so it is generally replaced $\widehat{\boldsymbol{D}}^{-\frac{1}{2}}\widehat{\boldsymbol{A}}\widehat{\boldsymbol{D}}^{-\frac{1}{2}}$, where $\widehat{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{I}_N$ and $\widehat{\boldsymbol{D}}$ are the degree matrices of $\widehat{\boldsymbol{A}}$ (Yan et al., 2021).

An example of graph convolution operation is shown in Figure 3. Figure 3A is form of convolution, and the right side shows the convolution on non-Euclidean space. Figure 3B shows the spectral convolutional operation. Given a 6-bus grid, the adjacency matrix $\boldsymbol{A}$ and degree matrix $\boldsymbol{D}$ are obtained based on the grid. Then, we can get $\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}$. Given the input data $\boldsymbol{x}_t^c$ and convolution kernel $g_{\theta 1}$, the result of graph convolution $\boldsymbol{x}_t^c *_G \boldsymbol{z}$ can be obtained.

The representation of the convolutional results on the network is further enhanced by the activation function ReLU:

$$\boldsymbol{h}_{\mathrm{f}}=g_{\mathrm{r}}\left(\sum_m^{M-1}\theta_m T_m(\tilde{\boldsymbol{L}})x\right) \tag{9}$$

where g_{r} is the activation function ReLU and $\boldsymbol{h}_{\mathrm{f}}$ is the spatial convolution output.

3.3.2 Multi-graph construction

Different dependencies are implied between loads at different locations in space. The load relationship implied by the different interconnections and distances of electric loads in different regions is called neighborhood dependence, and the relationship implied by the different load similarities due to the different patterns of electricity use is called load correlation. In order to mine the proximity dependence and load correlation of electric loads at each location in space, a physical connectivity map, location distance map, and virtual correlation map are constructed.

1) Physical connection graph: The connection matrix is established based on the connection relationship between the lines where the power loads are located at each location, i.e., the interconnections of the nodes in the grid topology. The element of this matrix can be defined as Eq. 10.

$$a_{i,j}^{\mathrm{a}}=\begin{cases}1, \text{node } i \text{ is connected to node } j\\ 0, \text{else}\end{cases} \tag{10}$$

2) Positional distance graph: A distance matrix is created based on the distance of the nodes where each power load is located $\boldsymbol{A}^{\mathrm{d}}$. The element of this matrix can be defined as Eq. 11.

$$a_{i,j}^{\mathrm{d}}=\begin{cases}\exp\left(-\dfrac{d_{i,j}^2}{\varepsilon_{\mathrm{d}}^2}\right), & d_{i,j}\geq\rho_{\mathrm{d}}\\ 0, & d_{i,j}<\rho_{\mathrm{d}}\end{cases} \tag{11}$$

where $a_{i,j}^{\mathrm{d}}$ is the element in $\boldsymbol{A}^{\mathrm{d}}$; $d_{i,j}$ is the distance between node i and node j; ε_{d} is the matrix threshold parameter; ρ_{d} is the distance threshold parameter.

3) Virtual similarity graph: A similarity matrix is created based on the similarity between the electrical loads at each location $\boldsymbol{A}^{\mathrm{s}}$ (Shi J. et al., 2023). The element of this matrix can be defined as Eq. 12, and the similarity between the load of node i and the load of node j can be calculated by Eq. 13.

$$a_{i,j}^{\mathrm{s}}=\exp\left(-\rho_{\mathrm{s}}c_{i,j}^{\mathrm{d}}\right) \tag{12}$$

$$c_{i,j}^{\mathrm{d}}=\sqrt{\sum_{t=1}^{T}\left(x_{t,i}^{l}-x_{t,j}^{l}\right)^2} \tag{13}$$

where $a^{s}_{i,j}$ is the element in $\boldsymbol{A}^{s}$; ρ_{s} is the parameter controlling the decay rate; $c^{d}_{i,j}$ is the similarity between the load of node i and the load of node j; $x^{l}_{t,i}$ and $x^{l}_{t,j}$ are the loads of node i and node j at time t.

The matrices $\boldsymbol{A}^{a}$, $\boldsymbol{A}^{d}$ and $\boldsymbol{A}^{s}$ are used to obtain the corresponding $\boldsymbol{L}^{a}$, $\boldsymbol{L}^{d}$ and $\boldsymbol{L}^{s}$ by bringing them into the standard computational form of the Laplace matrix, respectively. The Laplace matrices are brought into Eqs 8, 9 to obtain the convolution results of the loadings in each graph.

3.3.3 Multi-graph fusion

The graph fusion method is the key to graph neural networks, and a simple average summation of each graph will reduce the prediction performance. In this paper, we use the convolution results of each graph to be fused into a new graph by weighted summation to reflect the degree of influence of each graph in space. The weights of each graph are normalized using the Softmax function formulates as Eq. 14.

$$\boldsymbol{w}_{as}, \boldsymbol{w}_{ds}, \boldsymbol{w}_{ss} = g_{so}\left(\boldsymbol{w}_{a}, \boldsymbol{w}_{d}, \boldsymbol{w}_{s}\right) \tag{14}$$

where g_{so} is the Softmax function; $\boldsymbol{w}_{a}$, $\boldsymbol{w}_{d}$ and $\boldsymbol{w}_{s}$ are the learnable weight parameters of the physical connectivity graph, positional distance graph and virtual association graph, respectively; $\boldsymbol{w}_{as}$, $\boldsymbol{w}_{ds}$ and $\boldsymbol{w}_{ss}$ are the weights of the graphs after normalization, which indicate the influence degree of each graph in the new graph.

The weight parameters are multiplied with the results of the convolution of each graph and then summed, as shown in Eq. 15.

$$\boldsymbol{h}_{new} = \boldsymbol{w}_{as} \odot \boldsymbol{h}^{a}_{f} + \boldsymbol{w}_{ds} \odot \boldsymbol{h}^{d}_{f} + \boldsymbol{w}_{ss} \odot \boldsymbol{h}^{s}_{f} \tag{15}$$

where $\boldsymbol{h}_{new}$ is the convolution result of the new graph; $\boldsymbol{h}^{a}_{f}$, $\boldsymbol{h}^{d}_{f}$ and $\boldsymbol{h}^{s}_{f}$ are the convolution outputs of the physical connectivity graph, the positional distance graph and the virtual association graph, respectively.

The data features are extracted through the spatio-temporal convolution module, and the spatio-temporal convolution process is shown in Figure 4.

3.4 Output layer

The output layer is connected to the spatio-temporal prediction layer, which converges and transforms the passed results into the desired dimensions. The use of linear convolution can effectively transform the data dimension, and the selection of an appropriate activation function can extract the nonlinear features of the data. Due to the large degree of nonlinearity and high dimensionality of the data, this paper adopts the ReLU activation function and linear convolution twice in series, i.e., the predicted power load value is finally obtained without losing too much information each time. The prediction step size of this network is adjustable, i.e., the load value can be obtained at one time for more than one moment.

3.5 Spatio-temporal multi-graph prediction network structure

Before inputting the data, the resulting data should be blank-filled, outliers removed, and corrected (Azeem et al., 2021). The features such as historical electric load power, weather, and date are filtered using maximum information coefficient analysis to select the most relevant features as input data into the prediction network. The structure of the spatio-temporal multi-graph prediction network is shown in Figure 5, and the corresponding multi-graph convolution algorithm is shown in Algorithm 1.

Input: Data set of $\{\boldsymbol{X}_i\}_M$; the size of $\boldsymbol{X}_i$ is $i * f * t$; i is node, f is feature, and t is time.
Output: Multi-graph convolution model result $\widehat{\boldsymbol{Y}}$;

for each epoch **do**
  for each batch **do**
    Linear convolution: $Conv(X_i) \rightarrow X_{start}$;
    Initial value $0 \rightarrow X_{res}$;
    for each spatio-temporal convolutional layer **do**
      if first layer **then**
        $X_{start} \rightarrow X_{in}$
      else
        Previous X_{res} is current X_{in}: $X_{res} \rightarrow X_{in}$;
      end if
      Gated casual convolution: $Conv(X_{in}) \odot Conv(X_{in}) \rightarrow X_w$;
      Skip connection: $Y_{skip} + X_{skip} \rightarrow Y_{skip}$;
      Graph 1 convolution: $G1conv(X_{skip}) \rightarrow X_{G1}$;
      Graph 2 convolution: $G2conv(X_{skip}) \rightarrow X_{G2}$;
      Graph 3 convolution: $G3conv(X_{skip}) \rightarrow X_{G3}$;
      Graphs fusion: $X_{G1} + X_{G2} + X_{G3} \rightarrow X_{res}$;
      Residual connection: $X_{res} + Y_{res} \rightarrow Y_{res}$;
    end for
    Linear convolutions: $Conv(Conv(Y_{res})) \rightarrow Y_{out}$;
    Obtain MAE of network;
    Adjust hyperparameters;
  end for
end for
  Obtain result of $\widehat{\boldsymbol{Y}}$.

Algorithm 1. Spatio-temporal multi-graph prediction algorithm.

The prediction network mainly comprises a data embedding layer, a spatio-temporal prediction layer, and an output layer. The spatio-temporal prediction layer consists of multiple temporal and spatial convolutional blocks stacked together, enabling the network to capture data correlations at different temporal levels. Different spatio-temporal convolutional blocks converge different levels of information to the output layer through skip connections. In addition, residual connections are utilized in the blocks to accelerate convergence and to address possible degradation of the deep network (He et al., 2016). The overall flowchart is shown in Figure 6.

3.6 Evaluation indicators

The performance of the prediction network is evaluated by applying Mean Absolute Error (MAE) I_{MAE} calculated by Eq. 16, Mean Absolute Percentage Error (MAPE) I_{MAPE} calculated by Eq.

17, Mean Squared Error (MSE) I_{MSE} calculated by Eq. 18, and Root Mean Squared Error (RMSE) I_{RMSE} calculated by Eq. 19. MAE is the difference between the predicted load and the actual load, which truly reflects the prediction error, and in this paper, we choose the Mean Absolute Error as the loss function of the network.

$$I_{MAE} = \frac{1}{TM}\sum_{t=1}^{T}\sum_{m=1}^{M}\left|y_{t,m} - \hat{y}_{t,m}\right| \tag{16}$$

$$I_{MAPE} = \frac{1}{TM}\sum_{t=1}^{T}\sum_{m=1}^{M}\left|\frac{y_{t,m} - \hat{y}_{t,m}}{y_{t,m}}\right| \times 100\% \tag{17}$$

$$I_{MSE} = \frac{1}{TM}\sum_{t=1}^{T}\sum_{m=1}^{M}\left(y_{t,m} - \hat{y}_{t,m}\right)^2 \tag{18}$$

$$I_{RMSE} = \sqrt{\frac{1}{TM}\sum_{t=1}^{T}\sum_{m=1}^{M}\left(y_{t,m} - \hat{y}_{t,m}\right)^2} \tag{19}$$

where $y_{t,m}$ and $\hat{y}_{t,m}$ are the real and predicted values of load at time t node m respectively; m is the node number; and M is the total number of nodes.

4 Case study

4.1 Data set and parameters

In this paper, we use the 10 kV voltage level electric load dataset of a region in North China, including loads, weather conditions, date information and electricity prices, as shown in Table 2. All data have been desensitized and normalized to [0, 1]. The dataset contains a total of 10 bus data with a time range of 1 January 2020 to 1 June 2021, with a time interval of 60 min and a total of 24 points per day.

The predictive network model is implemented in Python software's PyTorch learning library. In the debugging process of the network, considering the size of the data volume, the data set is taken as 70% as the training set, 20% as the validation set, and 10% as the test set. After several comparative analyses and comprehensive prediction performance, the model parameters are set as shown in Table 2. Among them, the Dropout means to make the neurons not work in a certain proportion, which can make the model generalization ability stronger. In addition, the model is a single-step prediction, i.e., all bus loads at the next moment are predicted using all bus data of the previous day. The framework of the model of this paper is illustrated in Figure 7.

4.2 Feature selection results

The historical characteristics of Table 1 were analyzed by the maximum information coefficient analysis method to calculate the contribution of the influence of each characteristic quantity on the load, and the results are shown in Figure 8.

The meanings of the letter labels in the graph are shown in Table 1, and label I is the predicted day load. It can be seen that the MIC of historical load, time and electricity price with forecast daily load is high, which represents a strong correlation. And the MIC between electricity price and time is 1, which represents a high correlation with cyclical changes in time and electricity price on a daily basis. Due to the higher MIC between electricity price and load, electricity price was chosen as one of the input features. The MIC for weather conditions is generally between 0.2 and 0.4, with temperature and humidity having a greater impact on load. Too much input data will reduce the computing speed of the model, in order to have better performance of the prediction network, this paper takes the threshold of MIC as 0.3. In summary, the input features are historical load, temperature, humidity and electricity price.

4.3 Analysis of forecast results

The neighbor matrices in the physical connection graph, location distance graph and virtual association graph of the prediction model are shown in Figure 9. It can be seen that since some of the nodes are not directly connected to each other, the elements of the connection matrix are 0. To control the sparsity of the graph, the distance matrix elements of the two nodes that are too close to each other are set to 0 to improve the speed of operation. The similarity matrix elements vary as the nodes have different power usage patterns. We adjust the parameters ρ_s to make the distribution of the adjacency matrix more uniform. Applying the proposed spatio-temporal multi-graph prediction model, the metrics for evaluating power load forecasts at 10 nodes are shown in Table 3.As can be seen from Table 3, the mean absolute error varies from node to node due to their different load characteristics, and the overall MAE is 0.0136. The node 4 has a smaller MAE and a larger MAPE due to its small load power and high degree of fluctuation. However, node 6 and node 7 have irregular daily power loads with high uncertainty, resulting in larger MAPE and MSE. Node 3, node 5, node 8 and node 9 have smooth and distinctly cyclical load variations and have higher predicted MAPE. Node 1 has a higher load MAE and MAPE than node 5, but low MSE and RMSE, which indicates a higher degree of deviation from the individual results of the predictive model at node 5. Overall, the multigraph convolutional model predicted a MAPE of 5.26%.

4.4 Model comparison

To further validate the performance of the spatio-temporal multi-graph prediction network, it is compared with the following four widely used prediction networks:

1) Historical Average (HA): this model takes the average of the most recent load data as the predicted value and is one of the most classical statistical methods;
2) Gated Recurrent Unit (GRU): this network is a type of recurrent neural network that employs a gating mechanism to filter out the information in the long term sequences, thus improving the prediction performance;
3) Convolutional Neural Network-Long Short-term Memory Network (CNN-LSTM): this network utilizes a convolutional neural network to extract valid information from the input data. Due to the ability of LSTM to handle longer time series, they are integrated into the CNN for prediction;

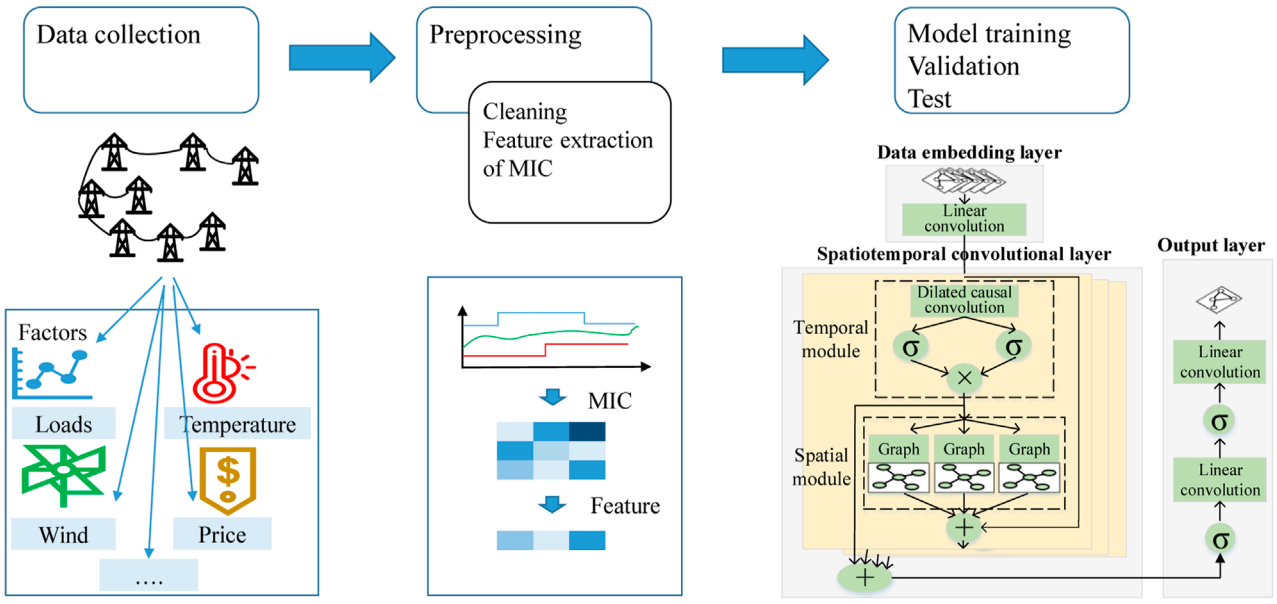


FIGURE 7
The framework of the model.

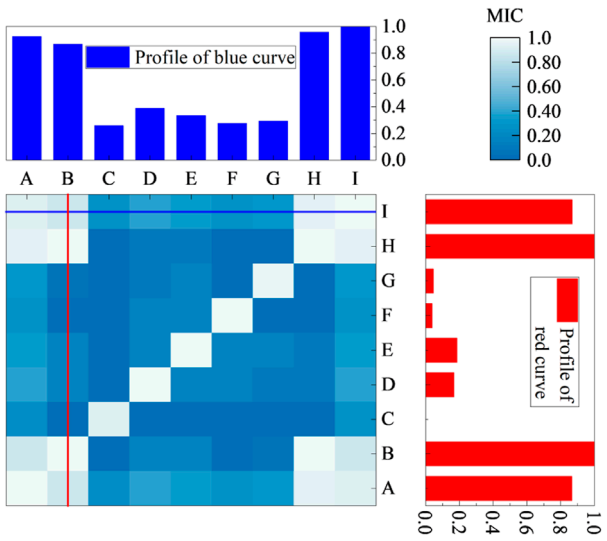


FIGURE 8
MIC of each feature.

4) Spatio-Temporal Convolutional Network (STGCN): this network consists of gated linear units to extract temporal features, graph convolutional networks to extract spatial features, and multiple spatio-temporal blocks superimposed to form a prediction network (Yu et al., 2017).

The prediction results are shown in Table 4. It can be seen that the performance of the multi-graph convolutional prediction model proposed in this paper are satisfactory. Since HA relies on simple averaging of historical loads to obtain the results, it is unable to capture the nonlinear factors of power loads in the time series, and thus has the lowest prediction accuracy. The GRU, a neural network with memory function, captures the correlation features of loads in the time series, with a MAPE of 6.62%, which reduces by 1.44% compared with that of HA, and achieves a better result. CNN-LSTM utilizes the convolutional network to process the feature information of the input load, which further reduces the MAPE by 0.31%.STGCN, a classical graph neural network, predicts a MAE of 0.0152 and a MAPE of 5.63%, which outperforms the traditional neural networks and statistical models. This is due to the fact that graph convolutional networks can process non-Euclidean load information and capture the hidden information of spatial loads. Due to the use of physical-virtual multi-graph structure to mine the different evolutionary relationships of loads in space, the proposed method has a MAPE of 5.26%, which is the best performance. The predicted MAE and MAPE evaluation metrics for each comparison method at each node load are shown in Figure 10.

It can be seen that the MAE and MAPE of each method are different due to the different fluctuation patterns of electric loads at each node. HA is a classical statistical model with large prediction errors in predicting more volatile loads such as nodes 4, 6 and 7. While deep learning models such as GRU and CNN-LSTM have less difference in MAPE at each node. The prediction networks with graph structure such as STGCN and MGCN can learn the potential relationship of each node and can further reduce the prediction error of each load.

4.5 Ablation experiments

In order to analyze the contribution of each module in the proposed physical virtual multi-graph network structure, we design ablation experiments. We compare the proposed model with the following variants:

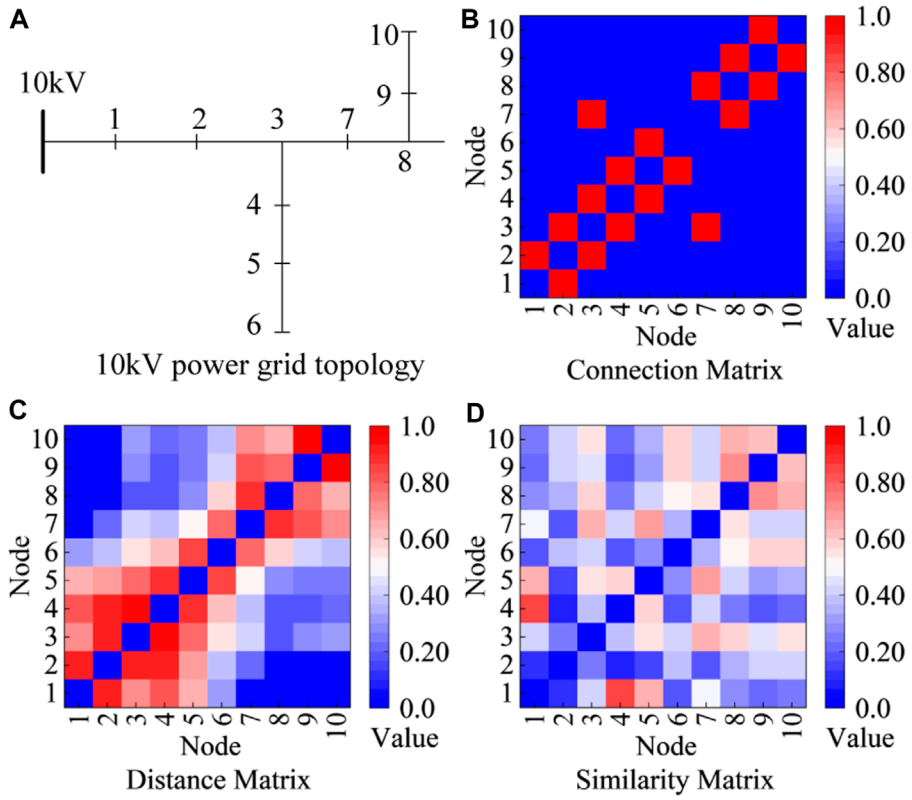


FIGURE 9
The adjacency matrices of the graphs: **(A)** 10 kV power grid topolpgy, **(B)** Convolution matrix, **(C)** Distance matrix, **(D)** Similarity matrix.

TABLE 3 Load prediction results of each node.

Node	MAE	MAPE/%	MSE/1e-5	RMSE
Node 1	0.00409	4.60	2.37	0.00487
Node 2	0.0219	4.09	75.34	0.0274
Node 3	0.00835	3.00	10.35	0.0101
Node 4	0.006794	9.61	7.52	0.00867
Node 5	0.00391	2.13	2.60	0.00510
Node 6	0.0359	8.13	194.23	0.0440
Node 7	0.0279	11.73	108.43	0.0329
Node 8	0.0101	2.93	15.54	0.0124
Node 9	0.000963	2.48	16.93	0.0130
Node 10	0.0169	3.87	63.41	0.0251
Average	0.0136	5.26	49.67	0.0222

TABLE 4 Load forecasting results of different methods.

Model	MAE	MAPE/%	MSE/1e-5	RMSE
HA	0.0225	8.06	104.84	0.0323
GRU	0.0192	6.62	86.58	0.0294
CNN-LSTM	0.0171	6.31	60.76	0.0246
STGCN	0.0152	5.63	52.79	0.0229
MGCN	0.0136	5.26	49.67	0.0222

- MGCN: The model is the proposed network, which contains the multi-graph and temporal convolutional network simultaneously.
- PC-GCN: In this variant, we retain the physical connection graph and remove the other graphs.
- P-GCN: Similarly, the virtual similarity graph is removed, retaining the positional graph and the physical connection graph.
- VS-GCN: This variant adopts virtual similarity matrix as features of graph, without employing the physical connection and distance graph.
- TCN: Different with above variants that contain the graph network, the variant is constructed with only temporal convolutional network.

The performances of different variants is shown in Table 5.

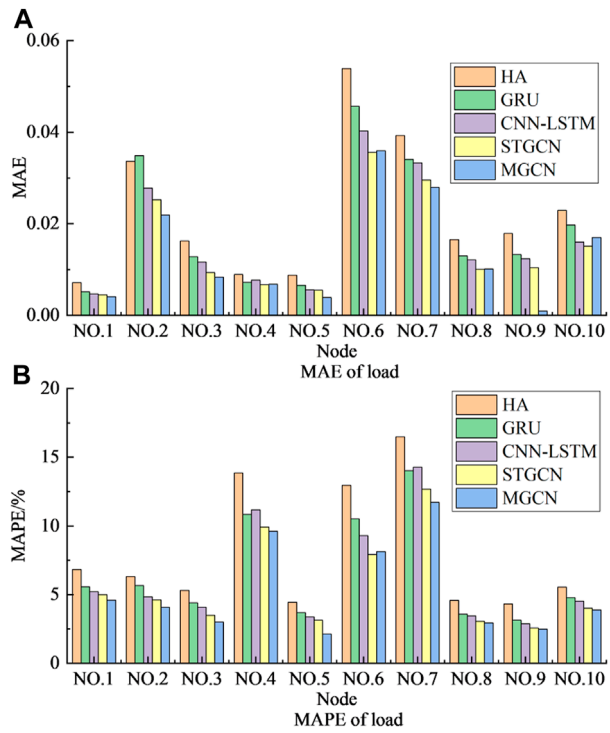


FIGURE 10
MAE and MAPE of each node: **(A)** MAE of load, **(B)** MAPE of load.

In Table 5, TCN obtains the MAPE of 6.40, which is similar to the performance of GRU and CNN-LSTM. Due to the lack of graph modules, they can only capture temporal features of data. When there is a physical graph in the model, the error in load prediction decreases significantly. In addition, We can observe that the MAE of P-GCN is very close to that of the STGCN. This indicates that physical graphs can extract hidden patterns of loads in the spatial dimension. We further combine a virtual similarity graph with the convolutional network, which achieves superior performance. Notably, the proposed virtual graph based on human domain knowledge can fully explore the evolution patterns of electricity loads.

5 Conclusion

In order to fully explore the correlation of various modes between the power loads of each node, this paper proposes a multi-graph convolutional spatio-temporal synergistic prediction method for power loads by fusing multi-dimensional information, and the theoretical analysis and the results of the arithmetic examples show that:

1) The maximum information coefficient method can effectively analyze load prediction influencing factors, select the most relevant features and reduce the redundancy of input information;
2) The non-Euclidean load information is processed by using spectral graph theory, and the constructed physical-virtual multi-graph convolutional network mines multiple spatial relationships between loads at each node, enriches the spatial characteristics of loads and improves the prediction accuracy;
3) Compared with statistical models, traditional neural networks and graph convolution models, the multi-graph spatio-temporal prediction network proposed in this paper has high prediction accuracy, which verifies the effectiveness of the method;

Although this paper has made some progress in constructing multi-graph convolution for spatio-temporal load prediction, the graph convolution network needs to be improved further: 1) The superior performance of the graph convolution network requires multiple rounds of manual hyperparameter tuning. More generalized and concise prediction networks can be considered for future adoption to improve the model's quality. 2) The electricity demand periodically changes over a large period. For example, the electricity load during the New Year is usually similar. We can add modules to learn load characteristics if there is continuous electricity data for every year. 3) We will design more general prediction models to achieve robust performance with incomplete data.

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

YJ: Supervision, Writing–original draft. WC: Conceptualization, Methodology, Software, Writing–original draft. XW: Software, Validation, Writing–original draft. YZ: Data curation,

TABLE 5 Load prediction results of different modules.

Indicators	MGCN	PC-GCN	P-GCN	VS-GCN	TCN
MAE	0.0136	0.0159	0.0153	0.0160	0.0178
MAPE/%	5.26	5.83	5.60	6.06	6.40
MSE/1e-5	49.67	54.88	52.52	56.48	68.32
RMSE	0.0222	0.0233	0.0229	0.0237	0.0252

Writing–review and editing. JZ: Validation, Writing–review and editing. WM: Software, Writing–review and editing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This work is supported by the Science and Technology Project of State Grid Shanxi Electric Power Company Limited (no. 5205E0220001).

Conflict of interest

Authors WC, XW, YJ, YZ, JZ, and WM were employed by State Grid Shanxi Electric Power Company Limited.

The authors declare that this study received funding from the Science and Technology Project of State Grid Shanxi Electric Power Company Limited. The funder had the following involvement in the study: conceptualization, data curation and formal analysis, visualization, writing the original draft and editing.

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

Aparicio, N., MacGill, I., Abbad, J. R., and Beltran, H. (2012). Comparison of wind energy support policy and electricity market design in europe, the United States, and Australia. *IEEE Trans. Sustain. Energy* 3, 809–818. doi:10.1109/tste.2012.2208771

Azeem, A., Ismail, I., Jameel, S. M., and Harindran, V. R. (2021). Electrical load forecasting models for different generation modalities: a review. *IEEE Access* 9, 142239–142263. doi:10.1109/access.2021.3120731

Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y. (2014). "Spectral networks and deep locally connected networks on graphs, 2nd int," in Conf. on Learning Representations, ICLR 2014—Conference Track Proceedings, Banff, AB, Canada, April 14-16, 2014.

Celebi, E., and Fuller, J. D. (2012). Time-of-use pricing in electricity markets under different market structures. *IEEE Trans. Power Syst.* 27, 1170–1181. doi:10.1109/tpwrs.2011.2180935

Cleary, B., Duffy, A., Oconnor, A., Conlon, M., and Fthenakis, V. (2015). Assessing the economic benefits of compressed air energy storage for mitigating wind curtailment. *IEEE Trans. Sustain. Energy* 6, 1021–1028. doi:10.1109/tste.2014.2376698

Deng, J., Wu, J., Zhang, S., Li, W., and Wang, Y.-G. (2024). Physical informed neural networks with soft and hard boundary constraints for solving advection-diffusion equations using fourier expansions. *Comput. Math. Appl.* 159, 60–75. doi:10.1016/j.camwa.2024.01.021

Fahim, S. R., Atat, R., Kececi, C., Takiddin, A., Ismail, M., Davis, K. R., et al. (2024). "Forecasting ev charging demand: a graph convolutional neural network-based approach," in 2024 4th International Conference on Smart Grid and Renewable Energy (SGRE), Doha, Qatar, 8-10 January 2024, 1–6.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, Seattle, WA, USA, June 14-19, 2020, 770–778.

Hou, H., Chen, Y., Liu, P., Xie, C., Huang, L., Zhang, R., et al. (2021). Multisource energy storage system optimal dispatch among electricity hydrogen and heat networks from the energy storage operator prospect. *IEEE Trans. Industry Appl.* 58, 2825–2835. doi:10.1109/tia.2021.3128499

Li, G., Li, Y., and Roozitalab, F. (2020). Midterm load forecasting: a multistep approach based on phase space reconstruction and support vector machine. *IEEE Syst. J.* 14, 4967–4977. doi:10.1109/jsyst.2019.2962971

Liao, W., Wang, S., Bak-Jensen, B., Pillai, J. R., Yang, Z., and Liu, K. (2023). Ultra-short-term interval prediction of wind power based on graph neural network and improved bootstrap technique. *J. Mod. Power Syst. Clean Energy* 11, 1100–1114. doi:10.35833/mpce.2022.000632

Lin, W., Wu, D., and Boulet, B. (2021). Spatial-temporal residential short-term load forecasting via graph neural networks. *IEEE Trans. Smart Grid* 12, 5373–5384. doi:10.1109/tsg.2021.3093515

Liu, L., Chen, J., Wu, H., Zhen, J., Li, G., and Lin, L. (2020). Physical-virtual collaboration modeling for intra-and inter-station metro ridership prediction. *IEEE Trans. Intelligent Transp. Syst.* 23, 3377–3391. doi:10.1109/tits.2020.3036057

Quilumba, F. L., Lee, W.-J., Huang, H., Wang, D. Y., and Szabados, R. L. (2014). Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. *IEEE Trans. smart grid* 6, 911–918. doi:10.1109/tsg.2014.2364233

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *science* 334, 1518–1524. doi:10.1126/science.1205438

Roy, A. D., and Yeafi, A. (2022). "Implementation of encoder-decoder based long short-term memory network for short-term electrical load forecasting," in 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI) (IEEE), Dhaka, Bangladesh, Dec. 17 2022 to Dec. 18 2022, 1–6.

Saber, A. Y., and Alam, A. R. (2017). "Short term load forecasting using multiple linear regression for big data," in 2017 IEEE symposium series on computational intelligence (SSCI) (IEEE), Honolulu, Hawaii, USA, 27 November - 1 December 2017, 1–6.

Samadianfard, S., Hashemi, S., Kargar, K., Izadyar, M., Mostafaeipour, A., Mosavi, A., et al. (2020). Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. *Energy Rep.* 6, 1147–1159. doi:10.1016/j.egyr.2020.05.001

Shi, J., Zhang, W., Bao, Y., Gao, D. W., and Wang, Z. (2023a). Load forecasting of electric vehicle charging stations: attention based spatiotemporal multi-graph convolutional networks. *IEEE Trans. Smart Grid* 15, 3016–3027. doi:10.1109/tsg.2023.3321116

Shi, P., Geng, L., Zhang, M., Xu, D., and Li, H. (2023b). "A short-term regional net load prediction method based on parallel fragment attention-bi-lstm," in 2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (IEEE), Shenyang, China, December 15, 2023 - December 17, 2023, 624–629.

Shi, W., Li, Y., Dong, S., Lu, X., Ye, H., and Hu, B. (2023c). "Short term load forecasting for holidays based on exponential smoothing of correlative correction," in 2023 International Conference on Power System Technology (PowerCon) (IEEE), Kuala Lumpur, Malaysia, 12-14 September 2022, 1–4.

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P. (2013). The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal Process. Mag.* 30, 83–98. doi:10.1109/msp.2012.2235192

Sun, S., Du, Z., Jin, K., Li, H., and Wang, S. (2023). Spatiotemporal wind power forecasting approach based on multi-factor extraction method and an indirect strategy. *Appl. Energy* 350, 121749. doi:10.1016/j.apenergy.2023.121749

Sun, S., Li, M., Wang, S., and Zhang, C. (2022). Multi-step ahead tourism demand forecasting: the perspective of the learning using privileged information paradigm. *Expert Syst. Appl.* 210, 118502. doi:10.1016/j.eswa.2022.118502

Tang, Z., Zhao, G., and Ouyang, T. (2021). Two-phase deep learning model for short-term wind direction forecasting. *Renew. Energy* 173, 1005–1016. doi:10.1016/j.renene.2021.04.041

Wang, Y., Rui, L., Ma, J., and jin, Q. (2023). A short-term residential load forecasting scheme based on the multiple correlation-temporal graph neural networks. *Appl. Soft Comput.* 146, 110629. doi:10.1016/j.asoc.2023.110629

Wei, C., Pi, D., Ping, M., and Zhang, H. (2023). Short-term load forecasting using spatial-temporal embedding graph neural network. *Electr. Power Syst. Res.* 225, 109873. doi:10.1016/j.epsr.2023.109873

Wu, Z., Pan, S., Long, G., Jiang, J., and Zhang, C. (2019). *Graph wavenet for deep spatial-temporal graph modeling.* arXiv preprint arXiv:1906.00121.

Xiu, C., Zhan, S., Pan, J., Peng, Q., Lin, Z., and Wong, S. (2024). Correlation-based feature selection and parallel spatiotemporal networks for efficient passenger flow forecasting in metro systems. *Transp. A Transp. Sci.*, 1–37. doi:10.1080/23249935.2024.2335244

Yan, B., Wang, G., Yu, J., Jin, X., and Zhang, H. (2021). Spatial-temporal Chebyshev graph neural network for traffic flow prediction in iot-based its. *IEEE Internet Things J.* 9, 9266–9279. doi:10.1109/jiot.2021.3105446

Yanmei, J., Mingsheng, L., Yangyang, L., Yaping, L., Jingyun, Z., Yifeng, L., et al. (2024). 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

Yu, B., Yin, H., and Zhu, Z. (2017). *Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting*. arXiv preprint arXiv:1709.04875.

Zhang, J., Liu, D., Li, Z., Han, X., Liu, H., Dong, C., et al. (2021a). Power prediction of a wind farm cluster based on spatiotemporal correlations. *Appl. Energy* 302, 117568. doi:10.1016/j.apenergy.2021.117568

Zhang, L., Xie, D., Luo, G., Qian, G., Song, M., and Chen, S. (2021b). "Research on short-term load forecasting based graph computation in power supply areas," in 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER) (IEEE), Jiaxian, China, July 27-31, 2021, 638–642.

Zheng, G., Chai, W. K., Zhang, J., and Katos, V. (2023). Vdgcnet: a novel network-wide virtual dynamic graph convolution neural network and transformer-based traffic prediction model. *Knowledge-Based Syst.* 275, 110676. doi:10.1016/j.knosys.2023.110676

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 22 July 2024
DOI 10.3389/fenrg.2024.1404165

Check for updates

OPEN ACCESS

EDITED BY
Yang Yu,
Nanjing University of Posts and Telecommunications, China

REVIEWED BY
Linfei Yin,
Guangxi University, China
Jinxing Che,
Nanchang Institute of Technology, China
Bowen Zhou,
Northeastern University, China

*CORRESPONDENCE
Xianfu Gong,
gongxianfu@gsbb.gd.csg.cn

RECEIVED 20 March 2024
ACCEPTED 12 June 2024
PUBLISHED 22 July 2024

CITATION
Peng B, Zuo J, Li Y, Gong X, Huan J and Liu R (2024), Short-term wind power prediction and uncertainty analysis based on VDM-TCN and EM-GMM.
Front. Energy Res. 12:1404165.
doi: 10.3389/fenrg.2024.1404165

COPYRIGHT
© 2024 Peng, Zuo, Li, Gong, Huan 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.

Short-term wind power prediction and uncertainty analysis based on VDM-TCN and EM-GMM

Bo Peng, Jing Zuo, Yaodong Li, Xianfu Gong*, Jiajia Huan and Ruoping Liu

Grid Planning and Research Center, Guangdong Power Grid Co., Ltd., Guangzhou, China

Due to the fluctuating and intermittent nature of wind energy, its prediction is uncertain. Hence, this paper suggests a method for predicting wind power in the short term and analyzing uncertainty using the VDM-TCN approach. This method first uses Variational Mode Decomposition (VDM) to process the data, and then utilizes the temporal characteristics of Temporal Convolutional Neural Network (TCN) to learn and predict the dataset after VDM processing. Through comparative experiments, we found that VDM-TCN performs the best in short-term wind power prediction. In wind power prediction for 4-h and 24-h horizons, the RMSE errors were 1.499% and 4.4518% respectively, demonstrating the superiority of VDM-TCN. Meanwhile, the Gaussian Mixture Model (GMM) can effectively quantify the uncertainty of wind power generation at different time scales.

KEYWORDS

wind power prediction, time sequence convolutional neural network (TCN), variational mode decomposition (VDM), Gaussian mixture model (GMM), uncertainty analysis

1 Introduction

Wind power, being the world's most significant new energy development focus, has seen global annual new wind power installations exceeding 50 GW since 2015. In 2019 alone, the newly installed capacity increased by 19% compared to 2018, reaching 60.4 GW (GWEC, 2022). Due to its inherent characteristics, accurate prediction of wind power is essential for grid-connected operations to ensure the smooth functioning of the power grid (Zhou et al., 2023). Hence, wind power prediction holds utmost importance.

In current wind power prediction research, the prediction time scale for wind power varies due to the impact of scheduling strategies (Zhou et al., 2023). In the long-term and medium-term prediction, wind power resources are generally predicted throughout the year to target wind power siting (Desalegn et al., 2023). At the same time, the installed capacity of wind farms is configured according to the range of prediction results. In addition, the results of short-term and ultrashort-term predictions over a 3-day period are usually used for bidding for feed-in services for wind power to guarantee power quality (Hong et al., 2019), and the results of wind power predictions are used for day-ahead or intraday scheduling of the grid (Jia et al., 2024).

In the realm of wind power prediction research, the prediction technology for wind power under various scenarios is commonly categorized into physical prediction and statistical prediction (Gu et al., 2021). Physical prediction studies typically involve the joint

simulation of atmospheric conditions (Zheng et al., 2022), and wind turbine behavior to forecast wind power generation over a specific time frame (Yang Y. et al., 2023). Statistical prediction techniques (Wang et al., 2024), which are typically applied to vast amounts of wind power data, utilize neural networks, multiple regression methods, and deep learning algorithms to model and predict wind power for future time periods (Meng et al., 2024).

As the depth of learning in intelligent algorithms continues to increase, the field of wind power prediction extensively employs these algorithms to develop novel prediction systems (Zhang et al., 2023). Within machine learning prediction models, temporal characteristics are commonly leveraged to assimilate historical wind power generation data (Sun and Zhao, 2020). Some researchers utilize historical and future forecast data from numerical weather predictions (NWP) to establish correlations between past inputs and outputs (Hong et al., 2019; Medina and Ajenjo, 2020). Wei et al. addressed the issue of low accuracy in ultra-short-term wind power prediction by proposing the use of LSTM for learning and prediction. They compared it with the traditional ARIMA model and found a significant improvement in prediction accuracy (Wei et al., 2023). Zhang et al. (2024) proposed a CNN-BiLSTM algorithm theory for multi-layer wind farm prediction, demonstrating a higher level of accuracy compared to traditional methods. While machine learning techniques have been widely applied in wind power prediction research, most studies have focused on their use for predictive purposes.

However, the accuracy of wind power prediction is influenced by various factors, and most individual algorithms are unable to address these challenges (Lin et al., 2024). Therefore, in recent years, hybrid algorithms have been commonly utilized in wind power prediction research (Zhu et al., 2023). For example, Yuan et al. introduced a hybrid model that combines the Least Squares Support Vector Machine (LSSVM) and the Gravitational Search Algorithm (GSA) for wind power generation prediction. They used the GSA algorithm to optimize the parameters of the LSSVM model in order to improve the prediction quality (Yuan et al., 2015). Zhou et al. (2019) proposed a K-Means-LSTM network model for wind power prediction and a bandwidth-optimised non-parametric kernel density estimation (KDE) model for probabilistic interval prediction of wind power. The K-Means clustering method is used to form different clusters of wind power impact factors to generate a new LSTM sub-prediction model. As well as non-parametric kernel density estimation generates intervals with narrower prediction intervals, higher interval coverage and higher prediction accuracy. Another study (Yuan et al., 2015) proposed a wind power prediction model based on the hybrid GWO-Copula approach to address the issue of wind power prediction distribution. It was observed that incorporating Copula with GWO (Grey Wolf optimization algorithm) significantly enhanced prediction accuracy without additional complexity. Additionally, Tu et al. developed an ARIMA-GARCH-T model to tackle the intricate timing challenges in wind power prediction, rectifying timing learning flaws and enhancing prediction accuracy (Tu et al., 2021). While the aforementioned research has made significant progress in optimizing wind power model parameters and improving model learning, there remains a limited focus on feature information processing.

In order to further enhance its ability of time series information extraction as well as anti-interference generalization, a combination of machine learning and load decomposition algorithms is often used (Zhang et al., 2018). To address the issue of poor model learning effectiveness, Deng D and colleagues developed a prediction method based on EEMD-GRU-MLR utilizing data characteristics. The Ensemble Empirical Mode Decomposition (EEMD) algorithm was employed for data decomposition, followed by evaluation of the prediction performance (Deng et al., 2020). EEMD serves as an enhanced version of Empirical Mode Decomposition (EMD). This technique necessitates the addition of white noise to the original signal to address spectral overlap, decay fluctuations, and trend information present in EMD. It filters out minor non-noise component fluctuations in the initial data, leading to irreversible loss of information. Consequently, the algorithm exhibits inherent limitations (Papazoglou et al., 2023).

Comparatively speaking, the signaling principle of the VMD algorithm is not complex, and the computational load is significantly smaller compared to EMD and EEMD. Moreover, its theoretical foundation is more robust. Unlike its predecessors, VMD does not rigidly define the meaning of each component but allows for independent selection of the number of components, enabling decomposition based on specific requirements (Kousar et al., 2022). However, new challenges have emerged with this algorithm. As each dimension of the data needs to be decomposed, predicted, and reconstructed separately, the computational time required remains substantial. To address issues related to limited algorithm accuracy, high computational complexity, lengthy model training times, low model generalization, and insufficient information extraction, this paper proposes a prediction method based on VDM-TCN for achieving high-precision wind power predictions.

Analyzing the uncertainty of wind power prediction is crucial. In uncertainty analysis methods, it can be divided into parametric methods and non-parametric methods. Parametric methods are based on point prediction models and assume the form of error distribution. However, this method may have limitations when dealing with diverse error distribution characteristics. In contrast, non-parametric methods use non-parametric estimation methods, do not need to assume the form of the target distribution, and can more accurately express the prediction error distribution, improving the analysis accuracy. For the uncertainty of wind power prediction, commonly used confidence interval methods are used for qualitative and quantitative analysis. The calculation of confidence intervals for uncertainty in wind power prediction can use parametric methods, non-parametric methods, and the decomposition and superposition of uncertainty factors. These methods help to better understand and address the uncertainty of prediction errors.

To ensure power grid stability, accurate assessment of future uncertainties in wind power bidding is crucial. While existing studies have delved into wind power prediction and uncertainty analysis, further exploration is needed to characterize multi-scale wind power prediction and uncertainty analysis. This study introduces a new wind power prediction framework based on VDM-TCN-EM-GMM to comprehensively investigate the relationship between the law and uncertainty of wind power prediction. By utilizing the VDM algorithm for data feature decomposition, the TCN algorithm for data prediction learning, and applying EM-GMM for qualitative and

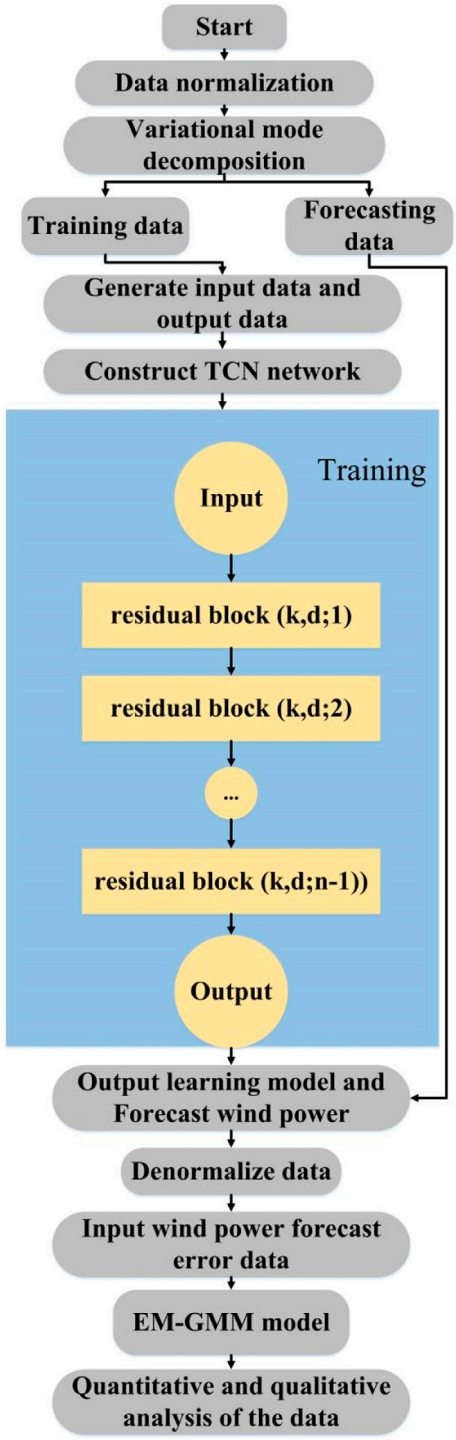


FIGURE 1
The VDM-TCN-EM-GMM technical process.

quantitative analysis, this study redefines the performance of wind power prediction uncertainty across multiple time scales, elucidating uncertainty patterns in wind power prediction at different time scales. The research aims to thoroughly examine uncertainties in wind power prediction, aiming to establish a robust prediction framework and provide valuable insights in this field. The process is shown in Figure 1.

(1) The key contributions of this study include the development of the comprehensive VDM-TCN-EM-GMM model, addressing challenges in quantitative wind power prediction and standardizing the process of wind power prediction uncertainty analysis. Compared to existing prediction algorithms and uncertainty analysis models, this framework can evaluate multi-time scale wind power prediction models comprehensively, enhancing the stability and accuracy of prediction results.
(2) Additionally, an in-depth investigation into prediction patterns and uncertainty characteristics across different time scales in various wind farms has been conducted, offering valuable data support and theoretical guidance for accurate wind power prediction in the future, bringing important insights for the development and application of the wind power industry.

Section 2 of this paper will introduce the principles and structures of the TCN model, the EM-based mixture Gaussian distribution model, and the confidence interval calculation model. Section 3 will present example analyses of the predictions for a wind farm using different models and time periods, along with uncertainty analysis and a comparison of the uncertainties in the confidence intervals. Finally, this study will be summarized in Section 4.

2 VDM-TCN model principle

The VDM-TCN model combines the advantages of variational mode decomposition (VDM) principles and time convolutional neural network (TCN) in a hybrid network. The VDM component decomposes the input wind power feature dataset into different modes, allowing the model to capture various fluctuation patterns present in the data. These modes are then fed into the TCN component, which utilizes temporal convolutional layers to learn the temporal dependencies and relationships of the wind power data features. The integration of VDM for mode decomposition and TCN for temporal modeling enhances the learning effectiveness of TCN, thereby improving prediction accuracy.

2.1 Principles of the TCN model

Temporal Convolutional Networks (TCNs) represent a neural network architecture specifically crafted for handling sequential data. TCNs employ one-dimensional convolutional layers to capture temporal relationships present in the input data. Through the utilization of dilated convolutions, TCNs can significantly enlarge the receptive field without a notable rise in the parameter count. This capability enables TCNs to effectively model extensive dependencies within the input sequence. Furthermore, TCNs integrate residual connections to aid in the training of deeper networks and address the issue of vanishing gradients (Yang S. et al., 2023).

In this model, the wind power feature dataset $X = \{x_1, x_2, x_3 \ldots\ldots x_t\}$ always corresponds to the wind power

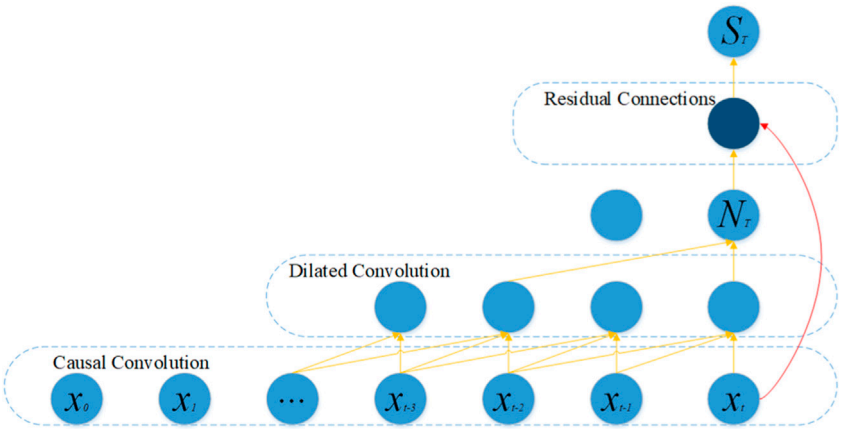


FIGURE 2
Structure of temporal convolutional network.

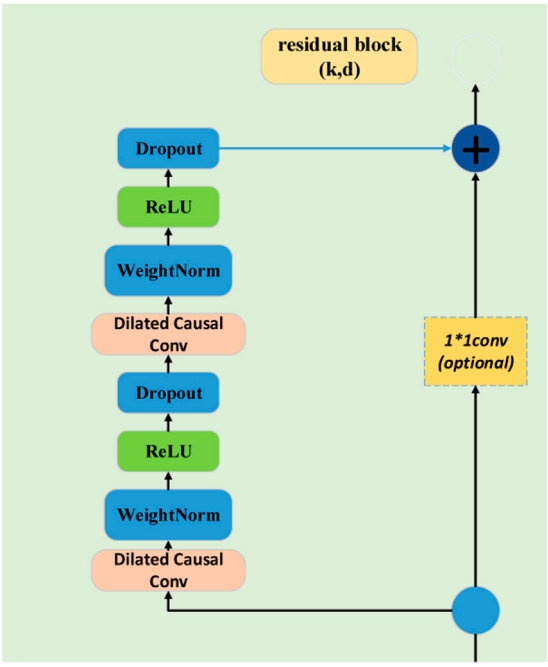


FIGURE 3
Diagram of residual block.

generation dataset $O = \{o_1, o_2, o_3.....o_t\}$. At the same time, an intermediate hidden layer $H = \{h_1, h_2, h_3.....h_t\}$ is introduced. All outputs satisfy the causal condition restriction, i.e., the current output y_t is only related to $\{x_1, x_2, x_3.....x_t\}$, and is not related to the "future" input $\{x_{t+1}, x_{t+2}, x_{t+3}.....x_{t+T}\}$. This is also in line with most of the time-series models in real life. This is consistent with most real-life applications, where future states are predicted with only historical data.

The relationship in the output can be represented as Eq. 1:

$$o_1, o_2, o_3.....o_T = f(x_1, x_2, x_3.....x_T) \quad (1)$$

What sets TCN apart from CNN models is that it incorporates both causal convolutions and dilated convolutions. Based on the actual data types and distributions, the network architecture of TCN for wind power prediction models in this study is depicted in Figure 2. The dimension of wind power input for $\{x_1, x_2, x_3.....x_t\}$ is 10, and output for $O = \{o_1, o_2, o_3.....o_t\}$ is 1.

2.1.1 Causal convolutional layer

In a time convolutional neural network, a causal convolutional layer ensures that each output element depends only on past input elements, In current wind power prediction model, this primarily refers to the correspondence between the wind speed, wind direction, temperature, and other data input at time t and the wind power generation at time t. This means that the layer does not have any connections to future input elements, preventing information leakage from the future. This property is crucial for tasks where the model should not have access to future information, such as in time series prediction or sequence modeling. The hollow causal convolution used in this paper combines the temporal constraints of causal convolution with the characteristics of dilated convolution in terms of skip sampling, ensuring that the output at the current time step depends only on the preceding states and is independent of the subsequent states (Guo et al., 2023). The formula for hollow causal convolution calculation is as follows Eq. 2.

$$f(x) = (o^* f)x = \sum_{t=0}^{k-1} f(i) X_{o-di} \quad (2)$$

Where X is the input, f is the filter, d and k are the dilation factor and convolution kernel size respectively. In the wind power prediction model based on the TCN algorithm in this study, the dataset's dilation factor is set to 8, and the convolution kernel is set to 20.

This paper incorporates dilated convolutional layers into the constructed TCN model. The expansion convolutional layer plays a crucial role in capturing complex patterns and relationships within the temporal data by increasing the richness of the learned representations. This process enables the network to extract more

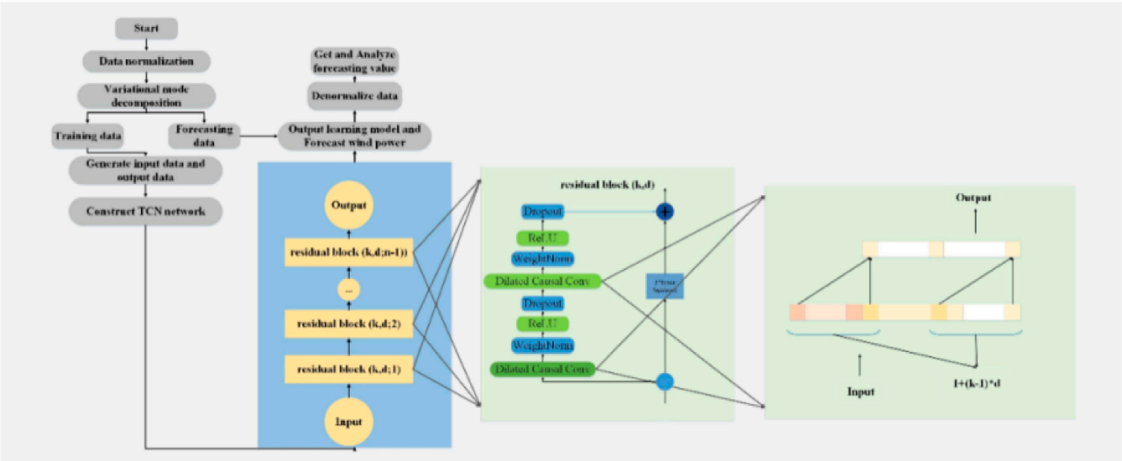


FIGURE 4
Wind power prediction process for VDM-TCN-EM-GMM.

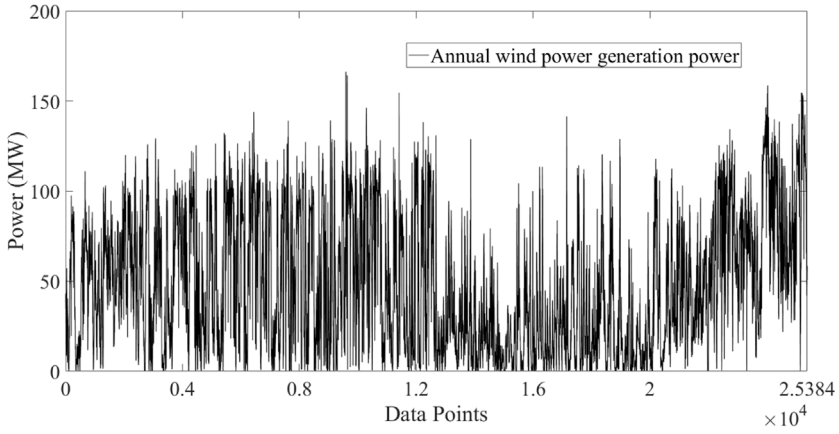


FIGURE 5
Annual wind farm power data.

intricate features and enhance its ability to learn and generalize from the input data.

2.1.2 Residual convolutional layer

Due to the large-scale data feature quantities and datasets required for wind power prediction training, using deeper networks can lead to the problem of gradient explosion. However, residual convolution can also improve the learning effectiveness of deep convolution. The residual convolutional layer in TCN (Temporal Convolutional Network) plays a crucial role in capturing long-range dependencies in sequential data. By incorporating residual connections, the network is able to learn the residual information between the input and output of each layer, allowing for easier optimization and training of deep networks. This enables the network to learn more effectively from the input data and improve the overall performance of the model. Furthermore, the residual convolutional layer in TCN allows for the efficient extraction of temporal features from sequential data by applying convolutional operations with shared weights across different time steps. This helps the network to capture complex patterns and dependencies in the data, leading to better generalization and prediction capabilities.

To address the channel width issue of wind power prediction data in matrices, the width of residual tensors is adjusted using 1*1 convolutions. As shown in Figure 3, in order to achieve complete coverage of the receptive field, residual blocks need to be incorporated into the TCN model. The width of the receptive field increases twofold with the addition of each residual block. Also, in order to avoid the saturation problem with multi-layer residuals, we increase the sparsity of the network by adding a corrected linear unit (ReLU) function to each layer of residuals. The calculation formula are as follows Eqs 3, 4:

$$r = 1 + \sum_{i=0}^{n-1} 2 \times (k-1) \times b^i \tag{3}$$

$$ReLU(x) = max(0, 1) \tag{4}$$

In this context, r represents the receptive field, k denotes the kernel size, and b stands for the dilation base.

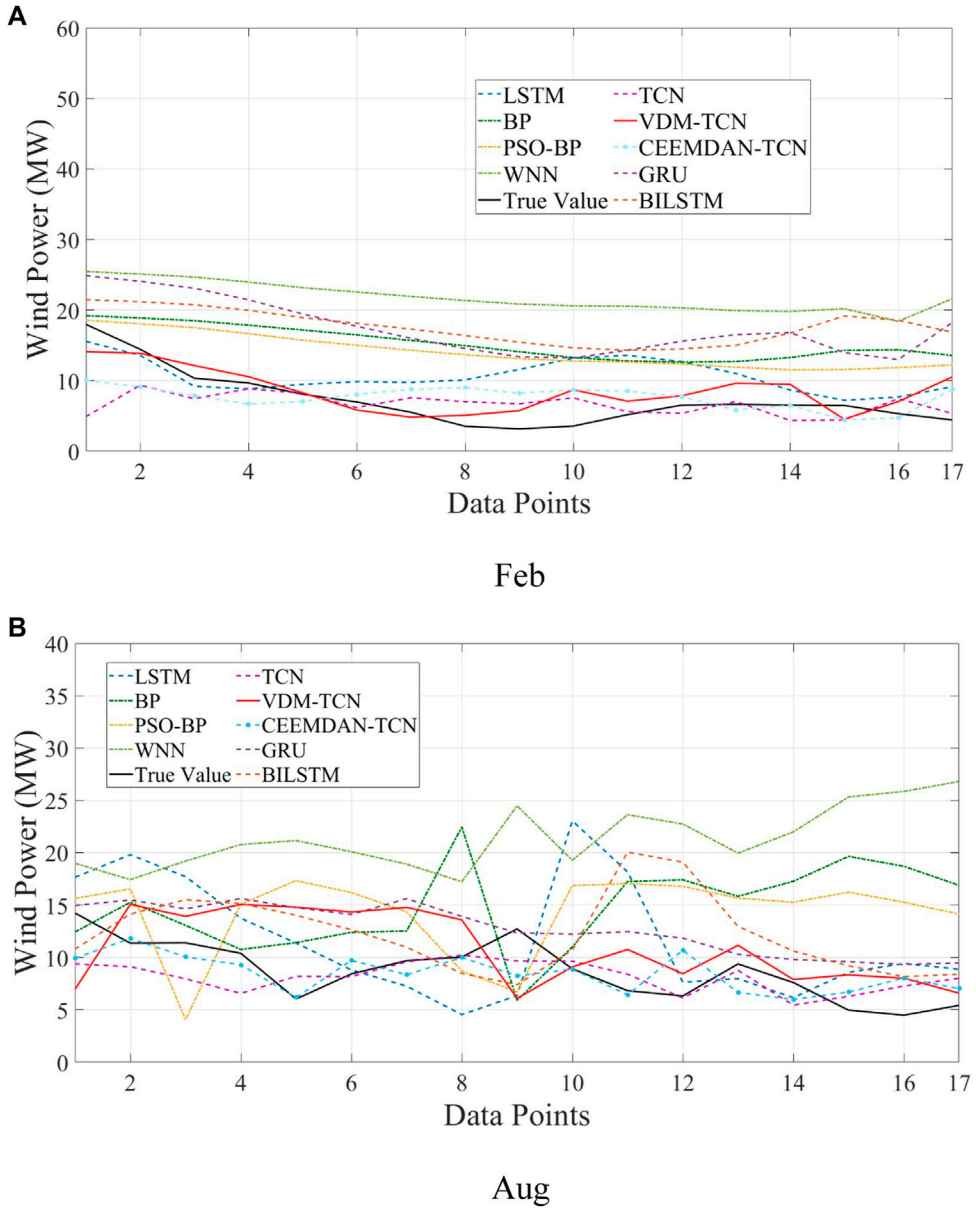


FIGURE 6
4-h wind power prediction results.

Finally, F(X) is added to X to obtain the output value y as in Eq. 5.

$$O = Activation(x + f(x)) \tag{5}$$

where $f(x)$ denotes the output of the convolutional layer and $Activation(.)$ denotes the activation function.

In terms of loss function design, this study uses the mean squared error (MSE) function to measure according to the actual characteristics of the training data as well as the specific network structure, and optimises the overall model by minimising the above error. The details are shown in Eq. 6.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - y_i^{'}\right)^2 \tag{6}$$

In this study, we found through experimental comparisons that RMSPropOptimizer can better ensure the stability of the error gradient of temporal convolutional neural networks during the training process, and it can modify the traditional gradient accumulation into an exponentially weighted moving average, so that it can adaptively regulate the change of the learning rate. Therefore, this study uses RMSPropOptimizer as an optimiser for TCN networks to better optimise the network model parameters. The formulas are given in Eqs 7, 8.

$$Sdw = \beta Sdw - (1-\beta)dw^2 \tag{7}$$

$$W = W - \alpha\frac{dw}{\sqrt{Sdw}} \tag{8}$$

where β is the smoothing constant, dw refers to the square of the gradient and W is the learnable parameter.

TABLE 1 Comparison of RMSE and MAE values for different prediction models in February.

			RMSE	SDE	MAE
		VDM-TCN	1.4999	1.3345	1.2225
		CEEMDAN-TCN	2.0777	0.5046	1.7029
		GRU	2.3859	2.2493	1.9842
		BILSTM	2.4608	2.3237	2.0019
		LSTM	2.7103	1.9744	2.1992
		TCN	2.2041	2.1749	1.4753
	4 h	PSO-BP	3.9827	1.3007	3.7643
		BP	4.6284	1.3607	4.4238
		WNN	6.1861	1.2952	6.049
		VDM-TCN	4.4518	4.2343	3.0803
		CEEMDAN-TCN	5.0996	4.9920	4.0605
		GRU	5.7955	5.7254	4.1027
		BILSTM	5.5582	5.5575	4.0593
		LSTM	5.9286	5.8766	4.3306
		TCN	5.3731	5.1953	4.1386
Feb	24 h	PSO-BP	7.2572	5.5628	6.5352
		BP	7.3778	5.7249	6.6567
		WNN	10.3357	5.1188	9.6935
		VDM-TCN	6.9254	6.9144	4.9209
		CEEMDAN-TCN	7.5615	7.2244	6.2156
		GRU	9.0590	7.8007	7.4141
		BILSTM	8.8224	7.1544	7.8590
		LSTM	9.7594	7.6034	8.176
		TCN	7.8642	7.8526	6.3630
	72 h	PSO-BP	10.2066	7.1924	8.483
		BP	10.4728	7.1958	8.8248
		WNN	17.6777	7.5986	16.0872

2.2 Variational modal decomposition

VDM (Variational Mode Decomposition) is a data-driven technique used for signal processing and analysis. It decomposes a signal into a set of modes that represent different oscillatory components of the signal. The process of VDM involves finding a set of modes that best capture the variations in the signal. This is achieved by formulating an optimization problem where the modes are obtained by minimizing a cost function that measures the differences between the original signal and its reconstructed version using the modes. The key idea behind VDM is to decompose the signal into a finite number of modes that are orthogonal to each other and capture different frequency components of the signal. This allows for a more efficient representation of the signal and can help in identifying and analyzing the underlying dynamics of the signal (Zhao et al., 2023).

Therefore, VDM needs to first use the Hilbert transform to calculate the analytic signal of each modal function $u_k(t)$, then mix the analytic signals of each mode with the central frequency $e^{-j\omega kt}$, and finally demodulate the signals using Gaussian smoothing and the gradient square criterion to obtain the bandwidth of each decomposition mode. The formula are as follows Eqs 9, 10:

$$\min\left\{\sum_{k=1}^{K}\left\|\beta_t\left[\left(\beta(t)+\frac{g}{\pi t}\right)\lambda_k(t)\right]e^{-gkwt}\right\|_2^2\right\} \tag{9}$$

$$s.t.\sum_{k=1}^{K}u_k = f \tag{10}$$

Furthermore, the optimization is enhanced by effectively solving through the utilization of penalty function α and Lagrange multiplier β.

$$L(\{u_k\},\{w_k\},\lambda) = \alpha\sum_{k=1}^{K}\left\|\beta_t\left[\left(\beta(t)+\frac{g}{\pi t}\right)\lambda_k(t)\right]e^{-gkwt}\right\|_2^2+\left\|f(t)-[\lambda_k(t)]\right\|_2^2 + \left[\gamma(t), f(t)-\sum_{k=1}^{K}\lambda_k(t)\right] \tag{11}$$

The alternating direction multiplier method is used in VMD to solve the variational problem of Eq. 11 by alternately updating ${u_k}^{n+1}$, ${\omega_k}^{n+1}$, and ${\lambda_k}^{n+1}$ to solve the improved Lagrangian expression "saddle point", i.e., the optimal solution of the constrained variational model in Eq. 9. where the modal components of the solution are new u_k and centre frequency ω_k, respectively:

$$\hat{u_k}^{n+1} = \frac{\hat{f}(\omega)-\sum_{i\neq k}\hat{u}_i(\omega)+\frac{\hat{u}_i(\omega)}{2}}{1+2\alpha(\omega-\omega_k)} \tag{12}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega|\hat{u_k}(\omega)|^2 d\omega}{\int_0^\infty|\hat{u_k}(\omega)|^2 d\omega} \tag{13}$$

2.3 EM-GMM model

As wind power prediction is affected by dataset errors as well as characterisation factors, a strong uncertainty is reflected in the prediction results, which is also known as error fluctuation. In order to ensure the competitive bidding of electricity and the stable operation of wind power, the uncertainty of wind power needs to be described to qualitatively and quantitatively analyse the prediction error of wind power to control the fluctuation range of the uncertainty error. Therefore establishing an error distribution for wind power prediction as well as establishing confidence intervals for wind power prediction is the best way to quantify the uncertainty of wind power prediction errors. The distribution models based on the combination of genetic algorithms and GMM are able to optimise these Gaussian mixture parameters using the selection, crossover and mutation operations of genetic algorithms to help the GMM fit the data distribution better. The advantage of this approach lies in its ability to fully leverage the global search capability of genetic algorithms while utilizing the flexible modeling capability of GMM, resulting in a more accurate description of the data distribution. Additionally, this combined approach can overcome the drawback of GMM being prone to local optima,

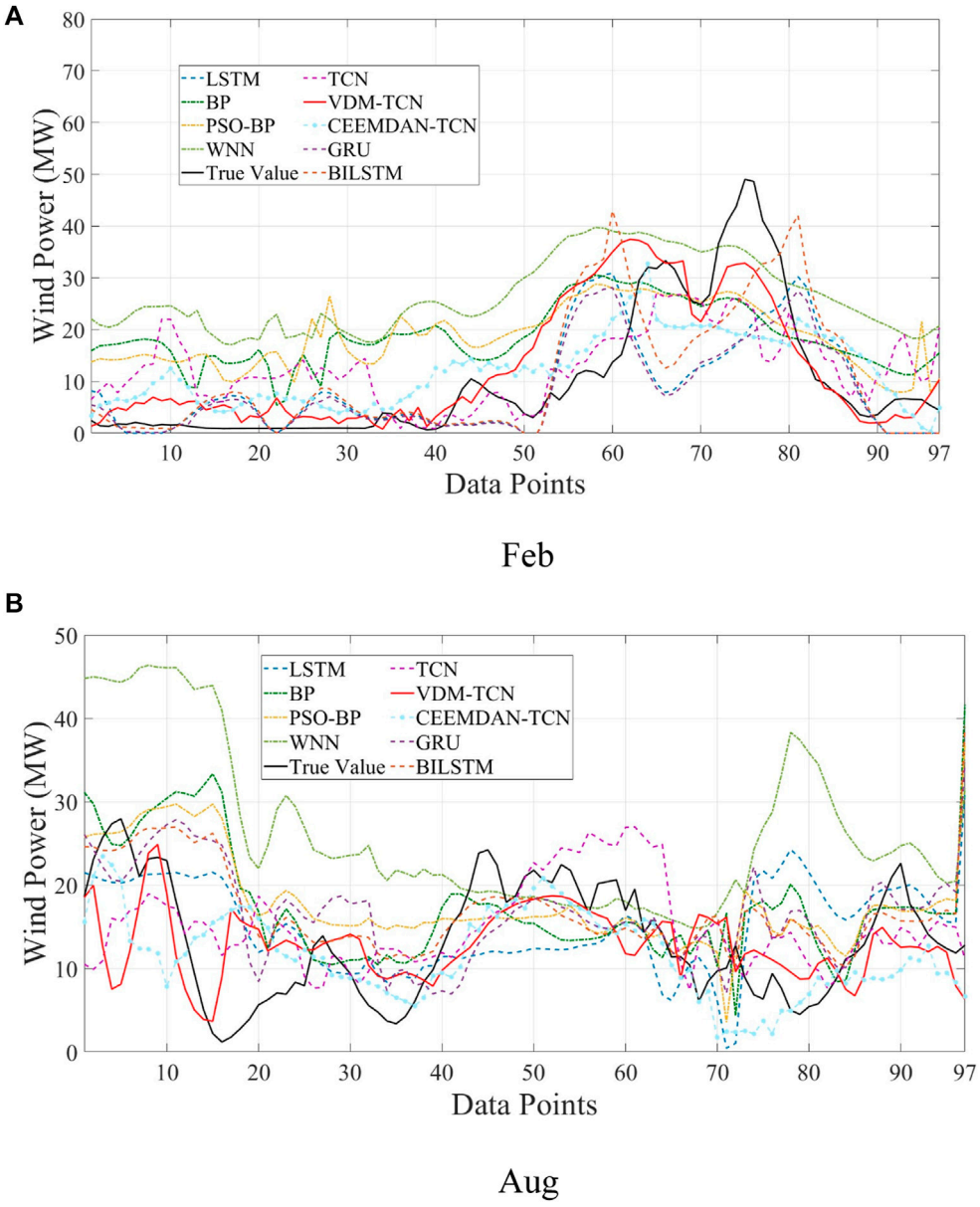


FIGURE 7
24-hours wind power prediction results.

thereby improving the robustness and generalization ability of the model.

2.3.1 Gaussian mixture models

GMM is a probabilistic model that represents a combination of multiple Gaussian distributions. Its structure is a method of approximating the probability distribution of a variable by linear mixing using a certain number of Gaussian functions.

Since each Gaussian component in the GMM mixture model is characterized by its mean and covariance matrix, these matrices determine the shape, position, and orientation of the distribution. At the same time, the data is generated by a mixture of Gaussian distributions under multiple weights and selections. Therefore, it is necessary to perform iterative training based on the EM algorithm, which estimates the parameters of the Gaussian components by maximizing the likelihood of the observed data, thus obtaining the optimal Gaussian parameter values. GMM is commonly used for clustering and density estimation tasks, aiming to divide the data into different groups based on the underlying distribution of the data. By using a combination of simple Gaussian components to capture the complex structure of the data, it is represented as follows Eqs 14, 15:

$$P(X_t) = \sum_{i=1}^{K} \omega_i \mu\left(P_t, \gamma_i, \sum_i\right) \tag{14}$$

$$\mu\left(P_t, \gamma_i, \sum_i\right) = \frac{1}{(2\pi)^{\frac{n}{2}} \left|\sum_i\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(P_t - \gamma_i)^T \sum_i^{-1} (P_t - \gamma_i)\right) \tag{15}$$

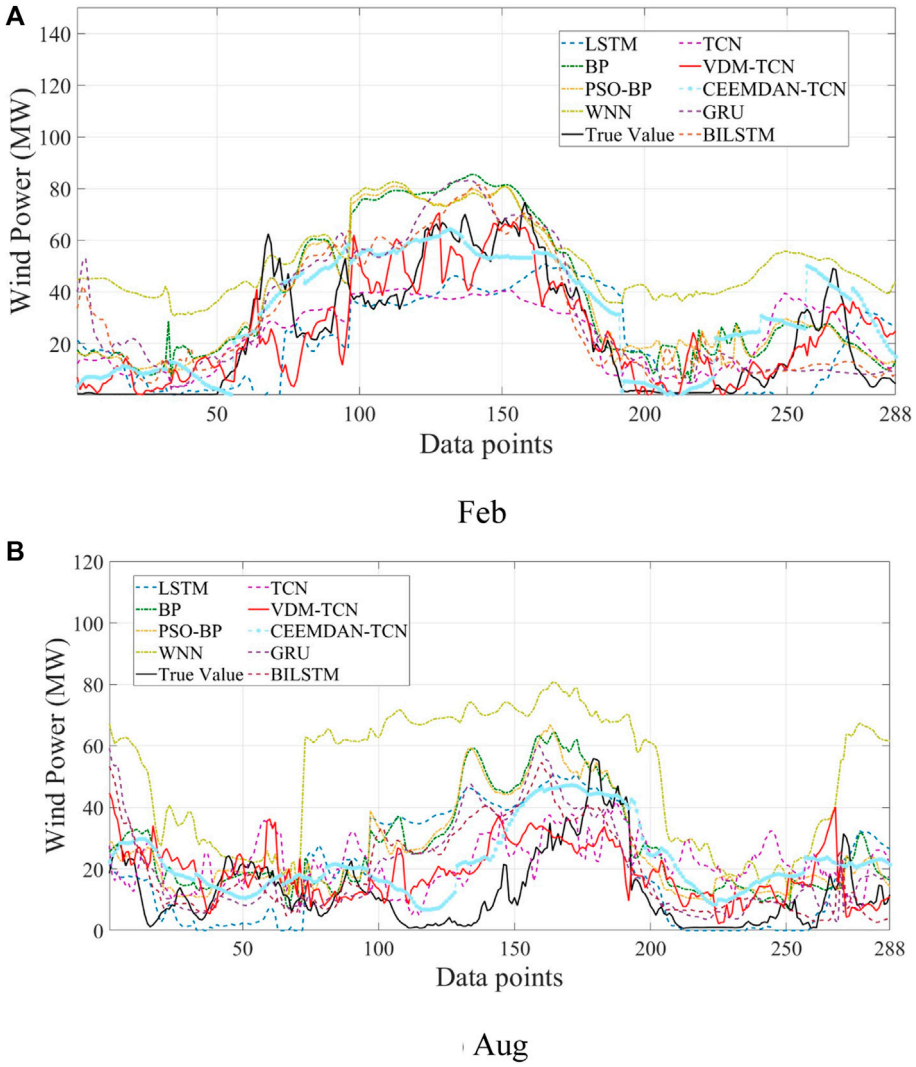


FIGURE 8
72-h wind power prediction results.

where n is the dimensionality of the pixel point in the high dimensional space, ω_i is the weight $\sum_{i=1}^{K}\omega_{i,t}=1$, μ_i and m are the mean and covariance matrices.

The parameter estimation of GMM is generally optimised by using the algorithm of EM for nonlinear probability functions during the training process, which greatly improves the implement ability of the algorithm under the premise of guaranteeing the accuracy. The specific principle is as follows.

Assuming $x_j=(\omega_j,\mu_j,\sum_j)$, $j=1,2,\cdots,K$, there are a total of K Gaussian models for the GMM, and all the parameters of the GMM are estimated through the sample set X: $\Theta=(x_1,x_2,\cdots,x_K)^T$, then the sample P is the log function with e as the base in Eq. 12, i.e., it can be written as ln, but the vast majority of representations of the log-likelihood function are still expressed in Eq. 16:

$$C(P\mid\Theta)=\log\prod_{m=1}^{T}H_K(P_i)=\sum_{m=1}^{T}\log\sum_{j=1}^{R}\omega_j\mu_j\left(P_i;\gamma_j,\sum\nolimits_j\right) \tag{16}$$

where T is the total number of samples, the parameters of the mixed model appropriate to the current sample set will maximise the log-likelihood function of Eq. 13, i.e., the estimation of the statistical parameters of the mixed model satisfies Eq. 17.

$$\Theta_0=\arg\max_{\theta}C(\Theta) \tag{17}$$

The EM algorithm is initially a statistical method that is an iterative algorithm. Assuming an initial estimate of the GMM parameters as $\Theta^{(0)}$, and assuming that the mixed model parameters for the q step iteration are $\Theta^{(q)}$, the $q+1$ step iteration process is:

(1) Calculate the expectation (E-Step)

Calculate the posterior probability that each data belongs to the $j-th$ class of distribution according to the parameters $\Theta^{(q)}$ of the current mixture model (Eq. 18):

TABLE 2 Comparison of RMSE and MAE values for different prediction models in August.

			RMSE	SDE	MAE
		VDM-TCN	1.0344	0.9949	0.3687
		CEEMDAN-TCN	1.2949	1.5188	0.9967
		GRU	2.5317	1.2042	2.2517
		BILSTM	3.2332	2.5696	2.6362
		LSTM	3.3593	2.9562	2.7157
		TCN	1.3011	1.2643	1.0480
	4 h	PSO-BP	4.3548	3.1014	4.0159
		BP	4.6529	3.2495	3.8920
		WNN	7.5868	2.8236	7.0418
		VDM-TCN	3.3869	3.3771	2.6583
		CEEMDAN-TCN	3.4552	3.3031	2.6473
		GRU	4.5569	4.1813	3.4885
		BILSTM	4.2636	3.7984	3.1175
		LSTM	4.6174	4.5279	3.7009
		TCN	3.5913	3.4681	2.9193
August	24 h	PSO-BP	4.9861	4.2973	3.7836
		BP	5.2140	4.7103	3.8039
		WNN	9.5918	6.3797	7.6712
		VDM-TCN	6.9690	6.0642	5.7049
		CEEMDAN-TCN	7.2858	6.6284	6.3456
		GRU	8.9160	7.8879	6.5391
		BILSTM	8.1683	7.2741	6.0462
		LSTM	9.9912	9.097	7.461
		TCN	7.9496	6.1311	6.7782
	72 h	PSO-BP	10.5957	7.6008	7.8764
		BP	11.0357	7.7612	8.3986
		WNN	23.2394	11.1917	20.5747

$$\varpi_{ij}^{(q+1)} = \frac{\omega_j^{(q)} \mu_j\left(P_j; \Theta^{(q)}\right)}{\sum_m^K \omega_r^{(q)} \mu_j\left(P_j; \Theta^{(q)}\right)};\ 1 \le m \le T, 1 \le j \le R \tag{18}$$

(2) Maximising expectation (M-Step)

After obtaining the posterior probability that each data belongs to each subclass, Eq. 14 is solved using gradient descent to obtain an estimate of Θ at step $q+1$.

Update the weights (Eq. 19):

$$\omega_j^{(a+1)} = \sum_{i=1}^{N} \varpi_{ij}^{(q+1)} \tag{19}$$

TABLE 3 MW level *t*-test distribution.

Time	Model	Error		RMSE	
		p	t	p	t
72 h	CEEMDAN-TCN	1.02E-07	−5.37	0.006	−3.12
	TCN	3.01E-14	−7.24	0.0021	−3.61
	BILSTM	3.02E-12	−7.01	0.0006	−4.73
	GRU	8.24E-09	−5.14	0.0002	−5.12
	LSTM	2.71E-12	−8.02	0.0007	−4.14
	PSO-BP	6.84E-23	−17.17	0.0003	−5.22
	BP	3.78E-23	−10.35	4.10E-05	−6.20
	WNN	7.77E-42	−12.91	3.53E-04	−4.21
24 h	CEEMDAN-TCN	0.0121	−1.24	0.0178	−1.93
	TCN	3.82E-04	−2.78	0.0104	−2.11
	BILSTM	1.12E-04	−4.55	0.0108	−2.45
	GRU	0.0187	−1.98	0.005	−3.56
	LSTM	3.21E-05	−4.12	0.0013	−3.69
	PSO-BP	4.10E-10	−6.57	3.56E-03	−4.58
	BP	1.24E-08	−4.77	4.42E-03	−4.61
	WNN	5.05E-27	−12.15	1.58E-04	−5.14
4 h	CEEMDAN-TCN	0.3212	−0.23	0.0121	−2.78
	TCN	0.2457	−0.48	0.0456	−2.54
	BILSTM	0.196	−0.68	0.0014	−3.45
	GRU	0.3257	−0.47	0.0031	−3.47
	LSTM	0.8796	−0.0023	2.98E-04	−4.23
	PSO-BP	0.02678	−1.57	2.12E-04	−5.74
	BP	0.23747	−1.18	1.47E-04	−5.78
	WNN	4.75E-04	−3.15	4.45–06	−7.01

Update mean values (Eq. 20):

$$\mu_j^{(q+1)} = \frac{\sum_{i=1}^{N} \varpi_{ij}^{(q+1)} X_i}{\sum_{i=1}^{N} \varpi_{ij}^{(q+1)}} \tag{20}$$

Update covariance matrix (Eq. 21):

$$\sum_j^{(q+1)} = \frac{\sum_{i=1}^{N} \varpi_{ij}^{(q+1)} \left(X_i - \mu_j^{(q+1)}\right)\left(X_i - \mu_j^{(q+1)}\right)^T}{\sum_{i=1}^{N} \varpi_{ij}^{(q+1)}} \tag{21}$$

Repeat steps (19), (20), (21) until $\|\Theta^{(q+1)} - \Theta^{(q)}\|$ sufficient hours to stop.

TABLE 4 Comparison of RMSE results for VDN-TCN models with different residual convolution and number of VDM decompositions.

	VDM^0-TCN^0	VDM^0-TCN^{10}	VDM^6-TCN^{10}	VDM^6-TCN^{20}
Count	10	10	10	10
Mean	10.98884	10.54142	10.26444	9.87834
Std	2.0306	1.9514	0.9786	0.4066
Min	10.77	10.1686	9.7089	9.385
25%	10.799475	10.3333	10.0377	9.70555
50%	10.8646	10.6576	10.3815	9.8457
75%	11.078975	10.716475	10.517675	10.02035
Max	11.5925	10.8061	10.5914	10.4822

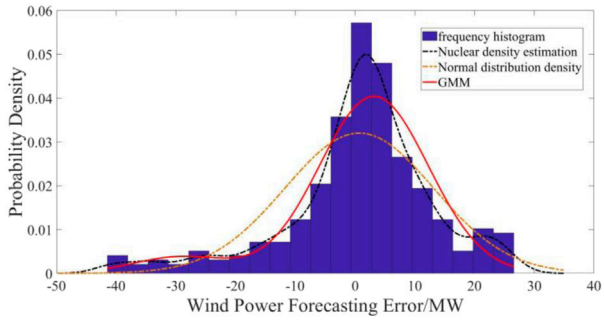


FIGURE 9
The 72-h probability density distribution of wind power prediction error under different distribution models.

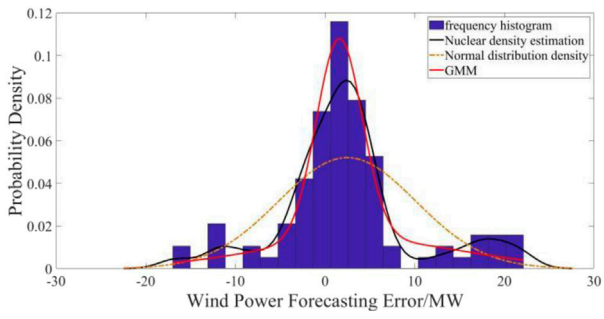


FIGURE 10
The 24-h probability density distribution of wind power prediction error under different distribution models.

2.3.2 Confidence intervals based on GMM

Based on the use of GMM, this study incorporates confidence interval calculation to quantitatively describe the uncertainty of predictions.

The wind power prediction error is the difference between the predicted value of wind power P_{fore} and the actual value of wind power P_{ture} at a certain point in time, as shown in Eq. 22.

$$e = P_{\text{fore}} - P_{\text{true}} \tag{22}$$

The formula is as follows:

$$P\left(\alpha_{\text{low}} < \alpha < \alpha_{\text{up}}\right) = 1 - \theta \tag{23}$$

In Eq. 23, $[\alpha_{\text{low}}, \alpha_{\text{up}}]$ is the refers to the upper and lower limits of the confidence interval. $1 - \theta$ is the reliability of the true value in the interval.

For uncertainty analysis modeling, it is challenging for overall error modeling or single-point error modeling to consistently demonstrate high reliability and adaptability at all times. Therefore, this study employs standard predictive analysis methods and Gaussian Mixture Model (GMM) for comprehensive analysis to enhance the clarity of predictions in uncertain scenarios.

The overall calculation steps are: Step1 Firstly, use the GMM method to establish the corresponding wind power error probability density map and calculate the wind power error probability density curve.

Step2 Under the given confidence level, find a shortest interval, so that the probability of the deterministic prediction error value falling into the interval is equal to the confidence level.

Step3 Use the (α_{up} and α_{low}) to derive the upper and lower limits of the wind power.

2.4 Data preprocessing techniques and predictive evaluation indicators

There are many factors affecting the prediction results in wind power prediction, among which the accuracy of the data and the size of the data volume often determine the prediction results, so it is necessary to carry out relevant preprocessing of historical data.

2.4.1 Wind power data screening

In the actual wind power generation process, due to turbine maintenance or shutdown, the power generated will be negative or zero value, as well as non-normal circumstances NWP value sudden change, such as the wind speed is greater than 40 m/s.

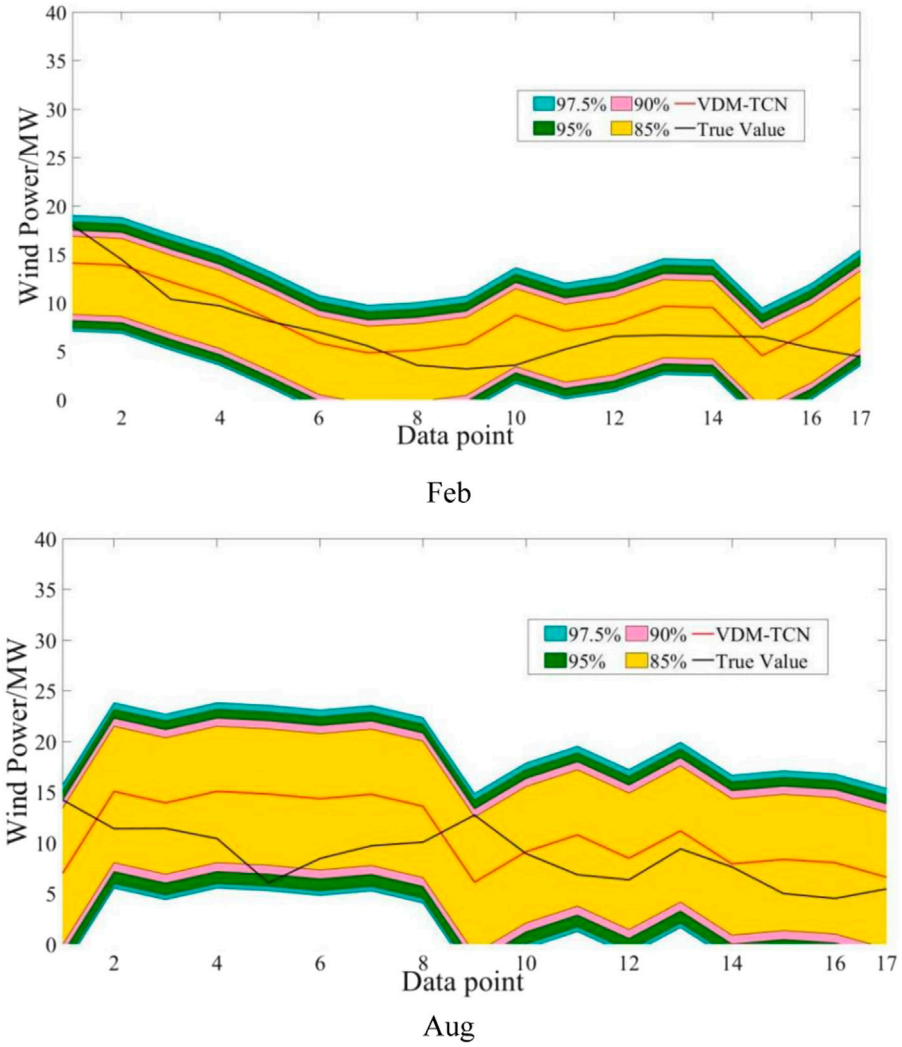


FIGURE 11
Distribution of confidence intervals for the 4-h prediction of the VDM-TCN model.

These data in the learning and prediction process will inevitably affect the learning effect, taking into account these factors, this paper in the data preprocessing of the data to be deleted.

2.4.2 Data standardisation

In order to improve the model's fitting results and reduce errors, the article conducted standardization processing with the following formula (Eq. 24):

$$x_{norm} = \frac{x - min}{max - min} \tag{24}$$

Where x_{norm} is the standard value of wind power; max indicates the maximum value of wind power data; min indicates the minimum value.

2.4.3 Evaluation index of deterministic prediction error

The root mean square error (RMSE) and mean absolute error (MAE) are used to evaluate the wind power forecast model. The formula are as follows Eqs 25–28.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(P_{\mathrm{true}} - P_{\mathrm{fore}})^2} \tag{25}$$

$$P_{\mathrm{RMSE}} = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(P_{\mathrm{ture}} - P_{\mathrm{fore}})^2}}{P_{\mathrm{cap}}} \tag{26}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}|P_{\mathrm{true}} - P_{\mathrm{fore}}| \tag{27}$$

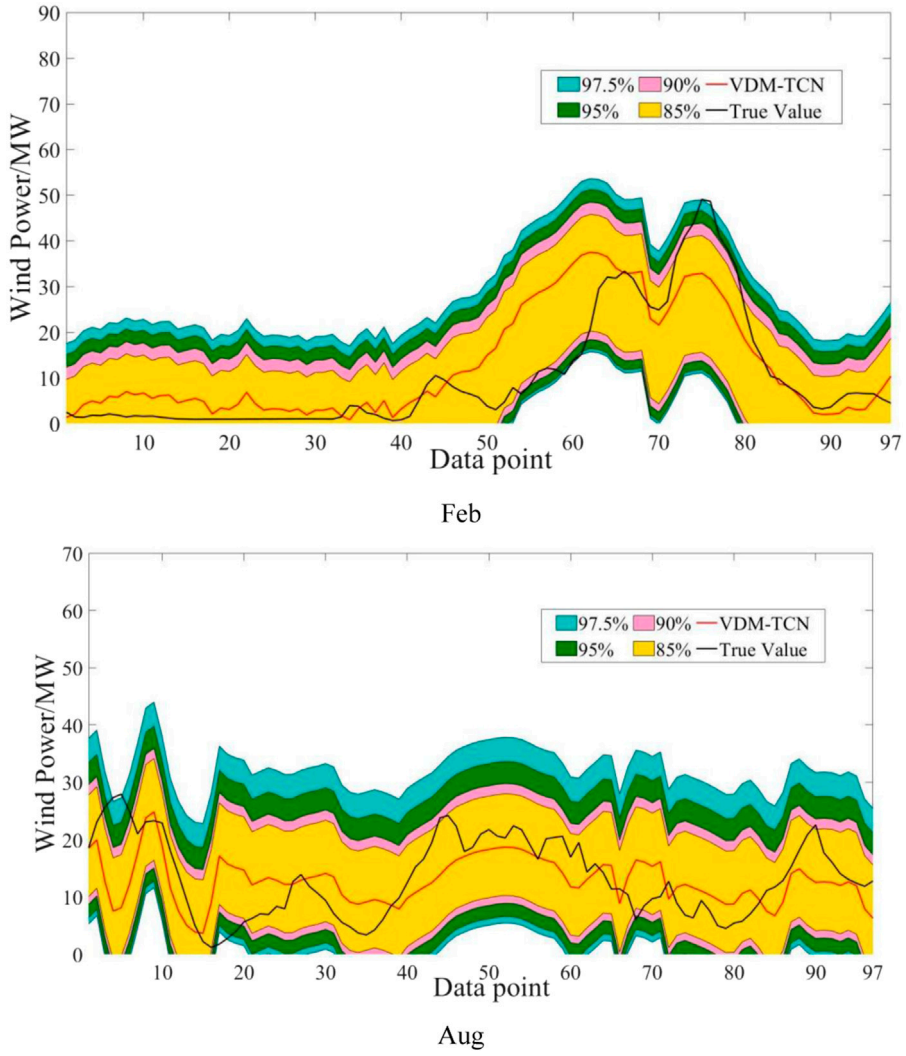


FIGURE 12
Distribution of confidence intervals for the 24-h prediction of the VDM-TCN model.

$$P_{\mathrm{MAE}} = \frac{\frac{1}{N}\sum_{t=1}^{N}|P_{\mathrm{true}} - P_{\mathrm{fore}}|}{P_{\mathrm{cap}}} \tag{28}$$

P_{cap} is the total installed capacity of the wind farm. P_{RMSE} and P_{MAE} is the ratio of the RMSE and MAE to the installed capacity.

2.4.4 Error evaluation indexes of uncertainty analysis methods

The coverage rate is used to evaluate the quality of the confidence intervals, as shown in Eq. 29.

$$\rho_p = \frac{1}{m} \times \sum_{i=1}^{m} \rho_i \tag{29}$$

where ρ_i is the coverage factor.

The technical route for short-term prediction and uncertainty analysis of wind power based on TCN-EM-GMM proposed in this paper is shown in Figure 4.

3 Case study

3.1 Data sources

The wind power data originates from a wind farm located in northern China, at 114°E longitude and 41°N latitude. The wind farm has an average elevation of 1,600 m and is equipped with 90 wind turbines, each with a power capacity of 1.5 MW. The entire dataset of a wind farm with a total installed capacity of 180 MW was chosen for prediction. The rotor diameter of the wind turbines is 70.5 m, and the tower height is 67 m. The wind power prediction data used in this study includes actual output power data from the wind farm's Supervisory Control and Data Acquisition (SCADA) system, as well as Numerical Weather Prediction (NWP) data for the wind farm. The time resolution of the actual output power data is 15 min. The NWP data is sourced from the National Meteorological Center, with a spatial resolution of 1 km. Therefore, there are multiple spatial grid

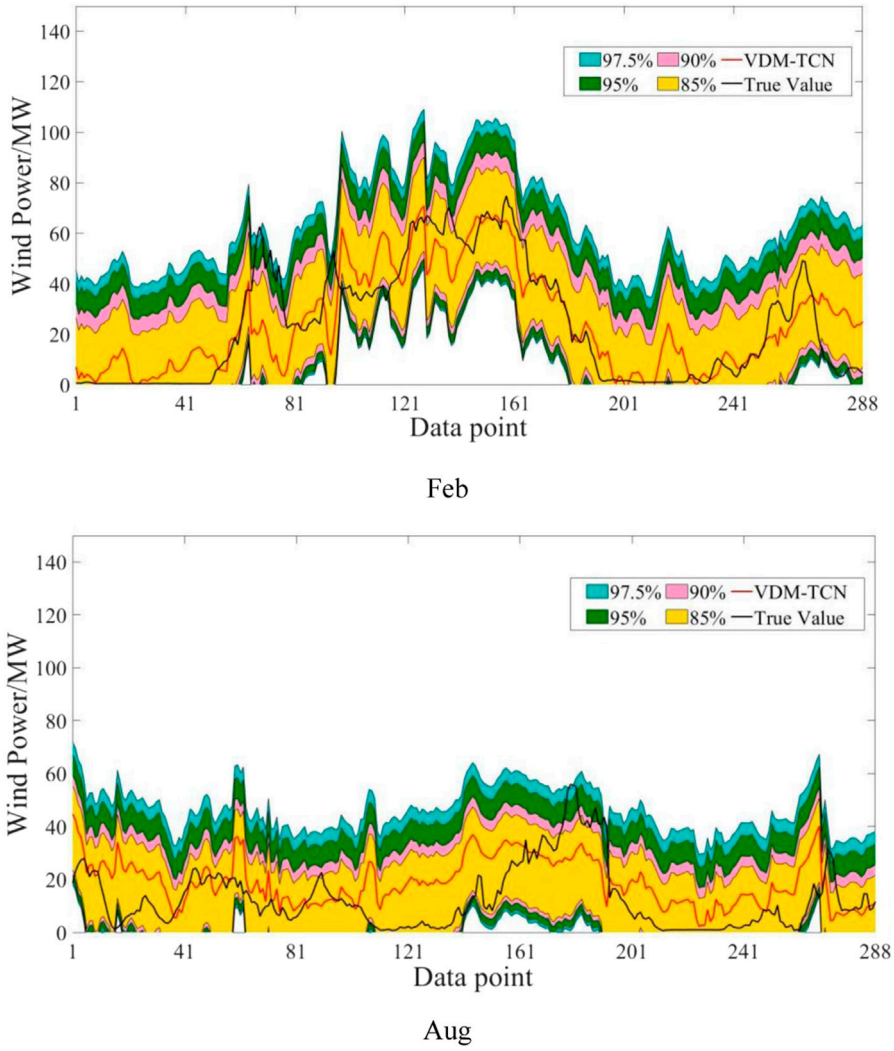


FIGURE 13
Distribution of confidence intervals for the 72-h prediction of the VDM-TCN model.

points with NWP data within the wind farm, and the average of these grid points' NWP data is used in this study. NWP data attributes include wind speed, wind direction, air pressure, temperature, and humidity, with a time resolution of 15 min. The experimental results are also in accordance with IEC standards.

As shown in Figure 5, we selected data from the entire year of 2010 and the first half of 2011 as the study data, with a time granularity of 15 min. To evaluate the effectiveness of the algorithm, we examined data from two specific intervals (February 10–13, 2011, and August 1–3, 2011) to understand their patterns across different time scales and seasons. The input data dimension is 10, which includes 6*n decomposed data groups from VDM-decomposed wind speed, and 1*n feature data groups for wind direction, air pressure, temperature, and humidity. The output dimension is 1*n wind power generation data, where n equals the number of data points at 15-min intervals required for training and prediction.

3.2 Wind power prediction and its uncertainty analysis

3.2.1 Wind power prediction analysis

The results of wind power generation prediction for 4-h intervals on February 4th and August 3rd in winter are presented in Figures 6A, B. The red solid line represents the VDM-TCN model. It can be observed that the VDM-TCN model aligns most closely with the actual values represented by the black solid line, followed by the CEEMDAN-TCN model. Furthermore, based on the values of 4-h RMSE and MAE displayed in Table 1, it is evident that learning conducted after VDM decomposition leads to a reduction in RMSE of over 0.8% compared to learning without decomposition. Additionally, the TCN model outperforms the LSTM, BP, PSO-BP, and WNN models in the 4-h prediction. The WNN model exhibits the poorest predictive performance, yet still remains within 7%. This can be attributed to the utilization of wavelet functions in

TABLE 5 Coverage rate of confidence interval for wind power based on VDM-TCN model.

month	Confidence level	72 h (%)	24 h (%)	4 h (%)
February	97.5	98.26	97.93	100
	95	96.53	95.87	100
	90	90.28	90.72	94.12
	85	86.46	85.54	88.24
August	97.5	98.26	97.94	100
	95	96.18	95.88	100
	90	92.71	91.76	94.12
	85	88.19	90.72	88.24

the activation function of the WNN, which results in suboptimal handling of the features of wind power data.

(Figures 7A, B) illustrates the 24-h wind power prediction outcomes for the 4th of February and 3rd of August during the winter season. The results indicate that VDM-TCN exhibits the highest prediction accuracy throughout the 24-h wind power prediction process. We observed a significant improvement in the prediction accuracy of the TCN algorithm after decomposing the wind power characteristic dataset, with an increase in RMSE of 0.3% compared to the undecomposed dataset. Additionally, the decomposition of the VDM algorithm exhibited higher adaptability than the CEEMDAN algorithm, showing superior predictive performance in February and August. From Figure 7A, it can be observed that, apart from VDM-TCN, the models exhibit significant discrepancies between the predicted results and the ground truth within the first 50 data points. This disparity can be attributed to the utilization of VDM for data decomposition, wherein certain anomalous frequency band data are extracted post decomposition. Consequently, the TCN model, by incorporating the anomalous characteristics of this data during the fitting process, achieves a more stable prediction outcome.

The forecast results for 72 h in Figure 8A, B show that in the 3-day forecast, VDM-TCN still maintains a significant advantage with better stability and more stable predicted values. TCN also demonstrates high predictive performance. However, data decomposed by VDM shows better learning and prediction compared to not using it. As shown in Figure 8B, although the learning and prediction effect of VDM on TCN is greatly sacrificed, some outliers still occur in small frequency bands. Through research, it was found that this is due to certain errors in the NWP values, and most importantly, the shutdown of some wind turbines in the wind farm due to wake effects and equipment damage reduces the matching degree between data and wind speed. This will also be a focus for future improvements.

Based on the comprehensive analysis of Figures 6–8 and Tables 1, 2, it is evident that VDM-TCN demonstrates superior predictive performance across various time scales. Additionally, the TCN model exhibits high stability during predictions, thus validating the effectiveness of the VDM-TCN model in wind power prediction. These findings provide data support for subsequent uncertainty analysis, with all RMSE prediction results falling within 8%.

The author employed a *t*-test to assess the significance of differences in prediction errors and RMSE results for May data from the same sample. A p-value ≤ 0.05 led to the rejection of the null hypothesis, indicating a significant difference in the predictive outcomes of the two models. Conversely, a p-value ≥ 0.05 resulted in the acceptance of the null hypothesis, suggesting no significant difference between the models' predictions. As presented in Table 3, there were statistically significant differences between the VDM-TCN model and other models in terms of 72-h and 24-h forecasts for both prediction error and RMSE, with negative t-values, indicating that the mean prediction error and RMSE of the VDM-TCN model were lower than those of the other models, thus confirming its superior predictive performance.

In order to better verify the accuracy and applicability of the model, we conducted ablation experiments for the VDN-TCN model with different residual convolution and number of VDM decompositions. Through Table 4, it can be clearly seen that the stability and accuracy of the prediction results are increasing with the addition of VDM and the introduction of residual convolution.

3.2.2 Wind power forecast and quantitative distribution analysis

Although the prediction errors of wind power generation can be qualitatively analyzed, it is still challenging to quantify them. In order to characterize the distribution of prediction errors in wind power generation quantitatively, this study utilizes GMM estimation to establish confidence intervals.

To calculate the confidence intervals for wind power generation prediction, the computation of probability density distribution is first required. In this study, a mixture of Gaussian model (GMM) and non-parametric kernel density estimation method are employed to obtain the probability density distribution of wind power prediction errors. Figures 9, 10 illustrate the probability distributions of wind power generation forecast errors for 72 h and 24 h. It can be observed that the non-parametric kernel density estimation method is more accurate than GMM in capturing trends across a wide range of distributions, but falls short in capturing certain abrupt changes at small scales compared to GMM. This discrepancy arises from the non-parametric kernel density estimation method's use of smoothing kernel functions to fit observed data points for modeling the true probability distribution curve, which is susceptible to bandwidth and data influences.

In order to better demonstrate the superiority of the GMM algorithm, we chose non-parametric kernel density estimation (NPKDE) and Gaussian modelling (GM) to contrast with the GMM algorithm and compare the uncertainty ranges of the predictions of different algorithms.

From the data, it can be observed that for Figures 11–13, the probability of the prediction values for the entire wind energy decreases with confidence intervals greater than the current confidence level. However, some forecasted values are not included in the confidence intervals due to actual output power changes caused by NWP errors, changes in operating states or gusts, and other factors. Furthermore, as the

confidence level increases, the width of the confidence interval also increases, with a higher probability of encompassing the forecasted values, which aligns with the principles of confidence interval calculations.

As shown in Table 5, we observe that the prediction intervals of the VDM-TCN model have high coverage rates at different confidence levels. Additionally, the VDM-TCN demonstrates high stability across various time ranges. Its coverage area meets the basic requirements of including the true values. This also proves that the GMM algorithm can accurately quantify the requirements of wind power prediction uncertainty.

4 Conclusion

This study innovatively proposes a method for short-term wind power prediction and uncertainty analysis using the VDM-TCN-GMM approach, which facilitates multi-scale short-term predictions of wind power via the VDM-TCN model. By applying variational mode decomposition technology to decompose the NWP, this method enhances feature diversity and improves data assimilation. Furthermore, the TCN model is utilized to identify and extract relationships among sequential features, thereby facilitating learning within a time-series framework. The Gaussian mixture model is also used to qualitatively analysis the uncertainty of wind power prediction and establish confidence intervals for quantitative analysis, and the following conclusions are drawn:

(1) The proposed VDM-TCN model not only has a temporal recursive nature, but also has an obvious advantage in feature extraction learning, which makes the VDM-TCN model have an obvious advantage in predicting wind power with time series characteristics.
(2) The prediction errors of the VDM-TCN model are all within 8%, with an improvement in RMSE prediction performance of over 1%.
(3) GMM is able to quantitatively calculate the distribution range and quantitative analysis of the prediction uncertainty in wind power generation. The coverage of the confidence interval is larger than the confidence level in 4 h, 24 h, and 72 h wind power prediction.

Although we have carried out multi-scale prediction and uncertainty analysis of wind power using VDM-TCN and EM-GMM algorithms, there is still a large amount of work that needs to be carried out for further research, and some of the much-needed work is as follows: 1) Wind power prediction needs to be further explored in terms of the impact of multi-source feature datasets on wind power prediction. 2) More algorithms need to be introduced into the field of wind power prediction to demonstrate the prediction performance of different models in different environments. 3) In terms of wind power uncertainty analysis, wind power uncertainty models will be further developed in the future to provide more accurate qualitative and quantitative analyses of wind power prediction.

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

BP: Conceptualization, Data curation, Formal Analysis, Methodology, Validation, Writing–original draft, Writing–review and editing. JZ: Data curation, Investigation, Project administration, Supervision, Writing–review and editing. YL: Project administration, Resources, Supervision, Writing–original draft. XG: Conceptualization, Data curation, Funding acquisition, Resources, Writing–review and editing. JH: Data curation, Investigation, Writing–review and editing. RL: Visualization, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Science and Technology Project of China Southern Power Grid Co., Ltd (No. 030000KC23040062 (GDKJXM20230367)).

Acknowledgments

We thank China Southern Power Grid Corporation Limited for providing financial and data support.

Conflict of interest

Authors BP, JZ, YL, XG, JH, and RL were employed by Guangdong Power Grid Co., Ltd.

The authors declare that this study received funding from the Science and Technology Project of China Southern Power Grid Co., Ltd. The funder had the following involvement in the study: manuscript preparation, data collection and analysis, and publication decisions.

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.2024.1404165/full#supplementary-material

References

Deng, D., Li, J., Zhang, Z., Teng, Y., and Huang, Q. (2020). Short-term electric load forecasting based on EEMD-GRU-MLR. *PST* 44, 593–602. doi:10.13335/j.1000-3673.pst.2019.0113

Desalegn, B., Gebeyehu, D., Tamrat, B., and Tadiwose, T. (2023). Wind energy-harvesting technologies and recent research progresses in wind farm control models. *Front. Energy Res.* 11. doi:10.3389/fenrg.2023.1124203

Gu, B., Zhang, T., Meng, H., and Zhang, J. (2021). Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation. *Renew. Energy.* 164, 687–708. doi:10.1016/j.renene.2020.09.087

Guo, G., Yuan, W., Liu, J., Lv, Y., and Liu, W. (2023). Traffic forecasting via dilated temporal convolution with peak-sensitive loss. *IEEE Intell. Transp. Syst. Mag.* 15, 48–57. doi:10.1109/mits.2021.3119869

GWEC (2022). *Global wind report 2022*. Brussels, Belgium: GWEC Europe Office.

Hong, D. Y., Ji, T. Y., Li, M. S., and Wu, Q. H. (2019). Ultra-short-term forecast of wind speed and wind power based on morphological high frequency filter and double similarity search algorithm. *Int. J. Electr. Power Energy Syst.* 104, 868–879. doi:10.1016/j.ijepes.2018.07.061

Jia, J., Zhang, G., Zhou, X., Shi, Z., Zhu, M., and Lv, X. (2024). Research on joint dispatch of wind, solar, hydro, and thermal power based on pumped storage power stations. *Front. Energy Res.* 12. doi:10.3389/fenrg.2024.1373588

Kousar, S., Zafar, N. A., Ali, T., Alkhammash, E. H., and Hadjouni, M. (2022). Formal modeling of IoT-based distribution management system for smart grids. *Sustainability* 14, 4499. doi:10.3390/su14084499

Lin, Q., Cai, H., Liu, H., Li, X., and Xiao, H. (2024). A novel ultra-short-term wind power prediction model jointly driven by multiple algorithm optimization and adaptive selection. *Energy* 288, 129724. doi:10.1016/j.energy.2023.129724

Medina, S. V., and Ajenjo, U. P. (2020). Performance improvement of artificial neural network model in short-term forecasting of wind farm power output. *J. Mod. Power Syst. Clean. Energy.* 8, 484–490. doi:10.35833/mpce.2018.000792

Meng, A., Zhang, H., Dai, Z., Xian, Z., Xiao, L., Rong, J., et al. (2024). An adaptive distribution-matched recurrent network for wind power prediction using time-series distribution period division. *Energy* 299, 131383. doi:10.1016/j.energy.2024.131383

Papazoglou, E. L. L., Karmiris-Obratanski, P., Karkalos, N. E. E., Thangaraj, M., and Markopoulos, A. P. P. (2023). Theoretical and experimental analysis of plasma radius expansion model in EDM: a comprehensive study. *Int. J. Adv. Manuf. Technol.* 126, 2429–2444. doi:10.1007/s00170-023-11292-6

Sun, Z., and Zhao, M. (2020). Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis. *Ieee Access* 8, 134422–134434. doi:10.1109/access.2020.3011060

Tu, Q., Miao, S., Yao, F., Li, Y., Yin, H., Han, J., et al. (2021). Forecasting scenario generation for multiple wind farms considering time-series characteristics and spatial-temporal correlation. *J. Mod. Power Syst. Clean. Energy.* 9, 837–848. doi:10.35833/mpce.2020.000935

Wang, Y., Zhao, K., Hao, Y., and Yao, Y. (2024). Short-term wind power prediction using a novel model based on butterfly optimization algorithm-variational mode decomposition-long short-term memory. *Appl. Energy.* 366, 123313. doi:10.1016/j.apenergy.2024.123313

Wei, J., Wu, X., Yang, T., and Jiao, R. (2023). Ultra-short-term forecasting of wind power based on multi-task learning and LSTM. *Int. J. Electr. Power Energy Syst.* 149, 109073. doi:10.1016/j.ijepes.2023.109073

Yang, S., Moreira, J., and Li, Z. (2023a). Bioinspired encoder-decoder recurrent neural network with attention for hydroprocessing unit modeling. *Ind. Eng. Chem. Res.* 62, 18526–18540. doi:10.1021/acs.iecr.3c01953

Yang, Y., Liu, J., Yang, Y., Xiao, J., and Alkhateeb, A. F. (2023b). An efficient hybrid method to predict wind speed based on linear regression and VMD. *Fractals* 31. doi:10.1142/s0218348x23401357

Yuan, X., Chen, C., Yuan, Y., Huang, Y., and Tan, Q. (2015). Short-term wind power prediction based on LSSVM-GSA model. *Energy Convers. manage.* 101, 393–401. doi:10.1016/j.enconman.2015.05.065

Zhang, J., Wei, Y.-M., Li, D., Tan, Z., and Zhou, J. (2018). Short term electricity load forecasting using a hybrid model. *Energy* 158, 774–781. doi:10.1016/j.energy.2018.06.012

Zhang, T., Huang, Y., Liao, H., Gong, X., and Peng, B. (2024). Short-term power forecasting and uncertainty analysis of wind farm at multiple time scales. *Ieee Access* 12, 25129–25145. doi:10.1109/access.2024.3365493

Zhang, T., Huang, Y., Liao, H., and Liang, Y. (2023). A hybrid electric vehicle load classification and forecasting approach based on GBDT algorithm and temporal convolutional network. *Appl. Energy.* 351, 121768. doi:10.1016/j.apenergy.2023.121768

Zhao, H., Zhang, H., Su, G., and Shi, X. (2023). Defect diagnosis method of cable shielding layer based on frequency domain reflection coefficient spectrum. *IEEE Trans. Electromagn. Compat.* 65, 114–125. doi:10.1109/temc.2022.3213351

Zheng, C., Yi, C., Shen, C., Yu, D., Wang, X., Wang, Y., et al. (2022). A positive climatic trend in the global offshore wind power. *Front. Energy Res.* 10. doi:10.3389/fenrg.2022.867642

Zhou, B., Ma, X., Luo, Y., and Yang, D. (2019). Wind power prediction based on LSTM networks and nonparametric kernel density estimation. *Ieee Access* 7, 165279–165292. doi:10.1109/access.2019.2952555

Zhou, K., Han, H., Li, J., Wang, Y., Tang, W., Han, F., et al. (2023). Interval model of a wind turbine power curve. *Front. Energy Res.* 11. doi:10.3389/fenrg.2023.1305612

Zhu, J., He, Y., and Gao, Z. (2023). Wind power interval and point prediction model using neural network based multi-objective optimization. *Energy* 283, 129079. doi:10.1016/j.energy.2023.129079

frontiers | Frontiers in Energy Research

TYPE Methods
PUBLISHED 11 October 2024
DOI 10.3389/fenrg.2024.1478834

Check for updates

OPEN ACCESS

EDITED BY
Jinran Wu,
Australian Catholic University, Australia

REVIEWED BY
Yongqiang Kang,
Lanzhou Jiaotong University, China
Zhesen Cui,
Changzhi University, China

*CORRESPONDENCE
Shiyao Hu,
841341970@qq.com

RECEIVED 11 August 2024
ACCEPTED 26 September 2024
PUBLISHED 11 October 2024

CITATION
Hu S, Rong C, Zhang M, Chai L, Ma Y and Zhang T (2024) Bad data identification method considering the on-load tap changer model. *Front. Energy Res.* 12:1478834. doi: 10.3389/fenrg.2024.1478834

COPYRIGHT
© 2024 Hu, Rong, Zhang, Chai, Ma 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.

Bad data identification method considering the on-load tap changer model

Shiyao Hu[1]*, Chunyan Rong[1], Mengnan Zhang[2], Linjie Chai[1], Yuxuan Ma[2] and Tianlei Zhang[2]

[1]Economic and Technology Research Institute of State Grid Hebei Electric Power Co., Ltd., Shijiazhuang, Hebei, China, [2]School of Electrical and Control Engineering, North China University of Technology, Beijing, China

With the connection and integration of renewable energy, the on-load tap-changer (OLTC) has become an important device for regulating voltage in distribution networks. However, due to non-smooth and non-linear characteristics of OLTC, traditional bad data identification and state estimation methods for transmission network are limited when applied to the distribution network. Therefore, the nonlinearity and droop control constraints of the OLTC model are considered in this paper. At the same time, the quadratic penalty function is introduced to realize the fast normalization of the tap position. It proposes a two-stage bad data identification method based on mixed-integer linear programming. In the first stage, suspicious measurements are identified using projection statistics and maximum normalized residual methods for preprocessing the measurement data. In the second stage, a linearization approach utilizing hyhrid data-physical driven is applied to handle nonlinear constraints, leading to the development of a bad data identification model based on mixed-integer linear programming. Finally, the proposed methodology is validated using a modified IEEE-33 node test feeder example, demonstrating the accuracy and efficiency of bad data identification.

KEYWORDS

bad data identification, on-load-tap-changer, hyhrid data-physical driven, mixed integer linear programming, linearization

1 Introduction

In the context of distribution network state estimation, the integrity of the estimation process can be compromised by the presence of erroneous data. Such data may arise from a variety of sources, including heterogeneous measurement instruments, sensor failures, and communication disruptions. These inaccuracies have the potential to severely impact the reliability and accuracy of the state estimation outcomes (Chen et al., 2021). The same time, With the continuous integration of renewable energy sources, OLTC is increasingly being utilized in power grid applications every year. It plays an important role in ensuring the safe and reliable operation of distribution networks and grid dispatching (Ju and Huang, 2023).The relevant parameters of the on-load tap-changer device exhibit nonlinearity and discreteness in mathematical models, and traditional continuous variable processing methods may reduce the accuracy of state estimation.

Traditional methods for bad data identification primarily encompass residual search methods (Handschin et al., 1975; Lin and Abur, 2018; Zhao and Mill, 2018), zero residual methods (Zhuang and Balasubramanian, 1987), and estimation-based methods (Huang and

Lin, 2004). These approaches are susceptible to errors such as misjudgment and missed detection. Contemporary techniques for bad data detection and identification predominantly include optimization-based and intelligent algorithm-based approaches. Among these, optimization-based methods have demonstrated a significant capacity for accurate identification of erroneous measurement data. A literature (Irving, 2008) proposes a robust state estimation model based on mixed integer nonlinear programming (MINLP). This method has high accuracy in identifying bad data, but lacks precision in estimation. Additionally, the model is nonconvex, nonlinear, and introduces a large number of 0/1 integer variables, making it difficult to solve. When the scale of the system increases, the solution efficiency decreases. The Schweppe-type generalized M-estimator with Huber psi-function (SHGM) is currently a method with good robustness (Mili et al., 1996). Introduces coefficients that reflect the lever-age properties, which can suppress the weight of bad data in leverage points and weaken the effect of bad data residuals in the objective function. However, it cannot completely filter out the impact of bad data. Existing methods need further improvement in terms of computational efficiency, accuracy, and ability to handle bad data.

For state estimation problems involving discrete variables, modelling and solving based on MINLP methods may have low convergence and computational efficiency. In order to avoid solving MINLP problems, in traditional state estimation, OLTC tap positions are usually treated as continuous variables (Shiroi and Hosseinie, 2008), which may lead to biases in the estimated results due to mismatch with the actual operating characteristics of OLTC. The literature (Korres et al., 2004; Teixeira et al., 1992; Handschin and Kliokys, 1995) presents, in addition to the traditional treatment of continuous variables, using rounding or sensitivity analysis to round the tap positions to integers. These integer values are then taken as known parameters and used in the state estimation problem to solve state variables such as voltage magnitude and phase angle. The literature (Maalouf et al., 2013) develops a mixed integer quadratic programming model with discrete variables, which can effectively address the tap position rounding issue with high accuracy, although the approach is more complex. Furthermore, the literature (Nanchian et al., 2017) applies a hybrid particle swarm optimization method to solve MINLP problems that involve discrete tap positions, but the algorithm has a long computation time and low efficiency.

To address the challenges of nonlinearity and low efficiency in bad data identification for OLTC discrete variables, this paper introduces a two-stage bad data identification method utilizing a positive curvature quadratic penalty function to facilitate rapid adjustment of OLTC tap positions. This method enhances solution efficiency while maintaining the accuracy of identification. The main contributions are as follows:

1) In order to improve the accuracy of bad data identification, an optimization-based method for bad data identification is proposed in this paper. This method can accurately and effectively identify the presence of bad data in the measurement data, demonstrating a good identification accuracy.
2) To cope with the issue of low efficiency in solving traditional MINLP models due to a large number of discrete and nonlinear constraints, this paper presents a two-stage bad data identification model based on mixed integer linear programming (MILP). In the first stage, all measurements are distinguished using projection statistics and maximum normalized residual methods, generating a reduced set of suspicious measurements. In the second stage, a linearization model based on hyhrid data-physical driven is proposed to linearize nonlinear constraints, leading to a MILP-based bad data identification method that significantly improves the accuracy and efficiency of bad data identification.

The organizational structure of this paper is as follows: Section 2 introduces the method of identifying leverage points and suspicious measurement bad data based on projection statistics and maximum normalized residual method, and obtains the reduction of suspicious measurement set; Section 3 introduces the traditional MINLP bad data identification model, and proposes a hyhrid data-physical driven linearization model considering OLTC constraints, and constructs a bad data identification model based on MILP; Section 4 illustrates the experimental at IEEE-33 node test feeder; Section 5 provides the conclusion.

2 Stage 1: reduction of suspicious measurement set

In this paper, the reduced suspicious measurement set mainly consists of two parts. The first part involves the identification of leverage points through the use of projection statistics, while the second part involves the identification of suspicious measurements in non-leverage points using the maximum normalized residual method. The detailed explanation of the two algorithm processes is provided below.

2.1 Identification of leveraged measurements using projection statistics

This subsection describes the mathematical model for identifying leverage points based on projection statistics. Firstly, the connotation of leverage points and their impact on state estimation are introduced. Currently, there are two types of definitions for leverage points. As shown in Figure 1, the first type is based on regression model analysis, while the second type is based on the analysis of diagonal elements of the hat matrix. The difference between the two methods is as follows: The first type constructs a factor space composed of the measurement Jacobian matrix and the measurement vector, obtaining the distribution of row vectors in each group of measurement Jacobian matrix and the measurement vector in the factor space, and identifies outliers as leverage points; the second type is based on residual sensitivity analysis, namely, determining whether the measurement residual increases significantly when there is a large measurement error in the system, and identifying measurements where the measurement error cannot be positively fed back to the measurement residual as leverage points.

The first method based on regression model analysis is introduced as follows. The first-order Taylor expansion is

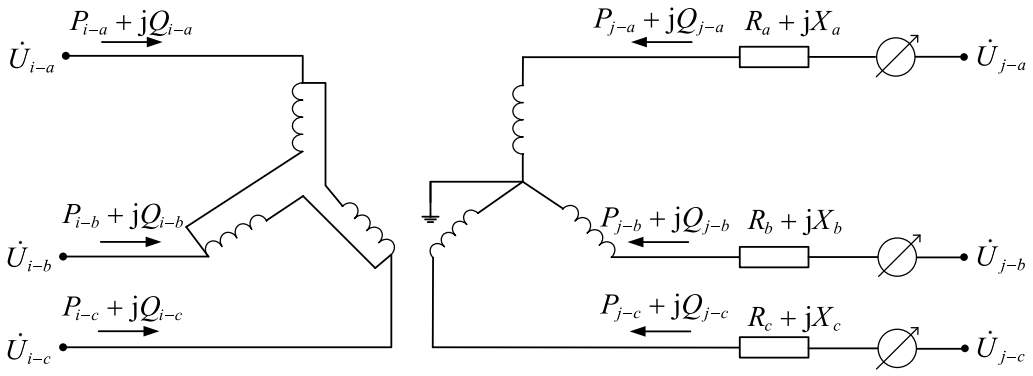


FIGURE 1
Dy11 on-load tap changer equivalent model.

performed at the initial value point x_0 and the approximate expression of the measurement error is obtained as follows:

$$\Delta z_i = H_i \Delta x + e_i \tag{1}$$

Where, Δz_i is the error between the true measurement value and the measurement vector; H_i represents the ith row element of the Jacobian matrix H; Δx represents the error between the current estimated value of the state variable and the initial value.

The above Equation 1 is called the regression analysis model in statistics. Δz_i is the output of the regression analysis model, Δx are the regression variables, and H is the coefficient matrix of the regression analysis. $(\Delta z_i, H_i)$ represents a point in the factor space and also indicates the relationship between the measurement vector and the true value, as well as the state variables. The regression factor is defined as, in the $m \times n$ dimensional Jacobi matrix, there is a total of m H_i,and the elements of each H_i correspond to an n-dimensional space coordinate, then they are all located in this n-dimensional space. These coordinates are called the corresponding measured factors, and the n-dimensional space is the factor space of the regression analysis.

Due to the system network parameters, measurement errors, and other reasons, abnormal values may appear in the above factor space, that is, the data is quite different from other data, and it is shown as outliers in the two-dimensional space. When outliers appear in the Y-space ' of the factor space, that is, in the ordinate axis ' Δz_i ' direction, they are bad data in the conventional sense of state estimation. When the outliers only appear in the ' X-space ' of the factor space, that is, in the ' h_i ' direction of the abscissa axis, the corresponding measurement is leverage measurement; when the outliers occur in ' X-space ' and ' Y-space ' at the same time, the corresponding measurement is the leverage measurement bad data. In the state estimation of power system, the leverage measurement is determined by the network topology, line parameters, measurement position and the meter error of the measurement instrument. Once the system network parameters are determined, whether the leverage measurement will become the inherent attribute of a certain measurement will not change with the change of the state variables and measurement values of the system.

To identify outliers in the "X-space," that is, to identify anomalies in the row vectors of the Jacobian matrix compared to other row vectors, this can be achieved by calculating distance measures between the individual row vectors. The Mahalanobis distance and other similar distance measures based on this can be used to calculate the distances between the row vectors. The criterion of such methods is to compare the distance between the row vectors with a set threshold. Measurements greater than this threshold are considered leverage measurements. The threshold setting criterion is to designate measurements that are far from most other measurements as leverage measurements. However, in cases where there are multiple leverage measurements in a system, problems may arise due to mutual masking between the leverage measurements, causing this type of method to potentially struggle to accurately identify systems with multiple leverage measurements.

The second method based on residual sensitivity analysis is introduced as follows. The estimation of the measurement error based on the least square method is defined as $\Delta\hat{z}$, and the matrix expression is as follows.

$$\begin{cases} \Delta\hat{z} = H\left(H^T R^{-1} H\right)^{-1} H^T R^{-1} \Delta z \\ W = H\left(H^T R^{-1} H\right)^{-1} H^T R^{-1} \end{cases} \tag{2}$$

Where, W is defined as a hat matrix.

When Equations 1, 2, the measurement residual is defined as the difference between the measured value and the estimated measurement vector, which can be equivalent to the error between the estimated measurement error value $\Delta\hat{z}$ and the measurement error value.

$$r = \Delta z - \Delta\hat{z} = (I - W)\Delta z = S\Delta z \tag{3}$$

In the equation, I is the identity matrix with diagonal elements equal to 1 and off-diagonal elements equal to 0; S defines the residual sensitivity matrix, with its matrix expression shown as shown in Equation 4:

$$S = I - H\left(H^T R^{-1} H\right)^{-1} H^T R^{-1} \tag{4}$$

The hat matrix W has the property of idempotence, where the matrix elements satisfy as shown in Equation 5:

$$W_{ii} = W_{ii}^2 + \sum_{i \neq j} W_{ij}^2 \tag{5}$$

From the above equation, Based on Equation 3, we know that when the diagonal elements of are close to 0. If the measurement error is large and the diagonal elements of the multiplication of the sensitivity matrix remains very small, indicating that the measurement error cannot be reflected in the residual of the measurement, then define this type of measurement as a leverage measurement. Furthermore, the leverage measurement can also be determined by comparing with its expected value to assess its relative magnitude.

The expected value is calculated as shown in Equation 6:

$$\bar{W} = E[K_{ii}] = \frac{1}{m}\sum_{i=1}^{m} K_{ii} = \frac{2n-1}{m} \tag{6}$$

If $W_{ii} \gg \bar{W}$,it is determined that the measurement corresponding to the diagonal element is determined to be a leverage measurement. On the basis of empirical experience, it is generally considered that when $W_{ii} \geq 2\bar{W}$, the measurement is a leverage measurement.

To avoid the difficult identification of leverage measurements that may arise from the two aforementioned methods, this paper adopts a method based on projection statistical values for the identification of leverage measurement. This method circumvents the calculation problem of the covariance matrix and directly utilizes the projections of the row vectors of the Jacobian matrix onto the relevant subspace to identify leverage measurements. First, calculate the projection statistical values for all measurements, with the calculation formula as follows:

$$PS_i = \max_k \frac{\left|H_i H_k^T\right|}{S_k} \tag{7}$$

Where PS_i represents the statistical projection value for measurement i, H_i represents the i-th element of the Jacobian matrix H; H_k^T represents the transpose of the k-th element of the Jacobian matrix H; S_k represents the covariance of H_i and H_k^T.

The calculation formula for S_k is given by the following equation:

$$S_k = 1.1926 \underset{i}{lomed}\ \underset{j \neq i}{lomed}\left|H_i H_k^T + H_j H_k^T\right| \tag{8}$$

Where, lomed is defined as $[(m+1)/2]$, $[x]$ represents the value of the integer part of n. For example,: $m = 6$, $[(m+1)/2] = 3$.

After calculating the projection statistical values corresponding to each measurement based on Equations 7, 8, compare them numerically with the projection statistical cutoff values. Under a Gaussian distribution, the projection statistics typically follow a Chi-square distribution, satisfying Equation 9:

$$b_i = \chi^2_{v,0.975} \tag{9}$$

Where, b_i is the calculated cutoff value.

When comparing the calculated projection statistical values with the cut off values, you can determine whether the measurement is a leverage measurement based on the comparison results.

$$D_i = \left(PS_i \cdot / b_i\right) > 1 \tag{10}$$

If the calculated projection statistical value D for measurement i satisfies Equation 10, then measurement i is determined to be a leverage measurement. The calculation of projection statistical values only involves simple algebraic operations and does not require matrix inversion. Therefore, even in cases where the Jacobian matrix is very sparse, accurate identification of leverage measurements can be achieved in a system with multiple leverage measurements.

2.2 Identification of suspicious bad data using maximum normalized residual method

As the calculation of residuals is approximate to the error values, it may not be able to detect bad data. By using standardized residuals, a more accurate method for identifying bad data can be obtained. The normalized residual value for measurement i can be obtained by dividing the residual value by the corresponding diagonal element in the residual covariance matrix.

$$r_i^N = \frac{|r_i|}{\sqrt{\Omega_{ii}}} = \frac{|r_i|}{\sqrt{R_{ii}S_{ii}}} \tag{11}$$

In Equation 11: r_i^N represents the normalized residual value; R represents the measurement error covariance matrix; S represents the sensitivity matrix; Ω represents the residual covariance matrix; r represents the residual value.

$$\Omega_{ii} = R_{ii} - h_i \cdot T_i \quad 1 \leq i \leq m \tag{12}$$

In Equation 12: Ω is the residual covariance matrix, $T = G^{-1}H^T$, G is the gain matrix, The calculation formula of G is $G = (H^T R^{-1} H)$, $R_{ii} = \sigma_i^2$ represent the measurement variance matrix of the error i.

When solving the weighted least squares state estimation, the residual value can be calculated. The calculation formula is as shown in Equation 13:

$$r = z - h(x) \tag{13}$$

For non-leverage measurement, the normalized residual vector r^N obeys the standard normal distribution, As Equation 14:

$$r_i^N \sim N(0,1) \tag{14}$$

Therefore, the maximum element in r^N is compared with the statistical threshold to determine the existence of suspicious measurement bad data. The threshold can be selected according to the required detection sensitivity level.

3 Stage 2: bad data identification based on mixed integer linear programming

In order to improve the solving efficiency of the traditional MINLP model, the hybrid data-physical-driven linearization model is first applied to linearize the non-linear constraints in the state estimation of on-load tap changer in transformers. Subsequently, an MILP-based bad data identification model is constructed. The following provides a detailed description of the

linearization process and the construction of the optimization model.

3.1 Bad data identification based on MINLP

The traditional MINLP-based bad data identification model is represented as follows:

$$\min \quad \sum_{i=1}^{m} b_i \tag{15}$$

$$s.t. \quad c(x) = 0 \tag{16}$$

$$h_i(x) - z_i^{\mathrm{m}} \geq -3\sigma_i - \mathrm{M}b_i \tag{17}$$

$$h_i(x) - z_i^{\mathrm{m}} \leq 3\sigma_i + \mathrm{M}b_i \tag{18}$$

In Equations 15–18: $c(x)$ represents equality constraints, including balanced node phase angle constraints, injection equality constraints, and load tap transformer droop control constraints., and $h_i(x)$ represents nonlinear measurement equations, including node voltage amplitude measurement equations, branch current amplitude square measurement equations, branch active power measurement equations, and branch reactive power measurement equations. z_i^{m} represents measurement values, M is a large positive number, taken as 10,000 in this paper. b_i is a binary variable. When b_i equals 0, it indicates that measurement i is not bad data, and the corresponding measurement equation inequality constraint is effective. When b_i equals 1, it can be determined that the measurement corresponding to i is bad data, and the corresponding measurement equation inequality constraint is invalid.

The bad data identification method based on mixed integer can overcome the problem that the residual method is difficult to identify the bad data of the multi-leverage measurement system, and has high identification accuracy. However, due to the serious non-convex nonlinearity of the model, the requirement for the solver is high, and the computational efficiency of the model solution is low.

3.2 The hybrid data-physical-driven measurement equation linearization model

3.2.1 OLTC model considering non-smooth control characteristics

OLTC includes two parts, "transformer" and "tap changer." Unlike traditional transformers, an OLTC sets the tap position as an unknown variable during modelling. By controlling the amplitude of the secondary side voltage to meet a certain dead band range, the OLTC adjusts the tap position to maintain the voltage level at the load center within a certain error range, thereby enhancing the power quality for the user. This paper proposes a droop control model for the OLTC, where the amplitude of the secondary side voltage and the tap turns ratio of the OLTC align with the droop control curve.

Define the column name of the nodal association matrix to represent the node $i-a$, $i-b$, $i-c$, $j-a$, $j-b$, $j-c$,the line name corresponds to the branch connected to $i-a$,the branch connected to $j-a$,the branch connected to $i-b$,the branch connected to $j-b$,the branch connected to $i-c$,the branch connected to $j-c$,where i、 j represent node; a、 b、 c represented by three phases of each node. The nodal correlation matrix is as shown in Equation 19:

$$C = branch \begin{bmatrix} \overset{node}{\begin{matrix} 1 & -1 & & & & \end{matrix}} \\ \begin{matrix} & & & 1 & & \end{matrix} \\ \begin{matrix} & 1 & -1 & & & \end{matrix} \\ \begin{matrix} & & & & 1 & \end{matrix} \\ \begin{matrix} -1 & & 1 & & & \end{matrix} \\ \begin{matrix} & & & & & 1 \end{matrix} \end{bmatrix} \tag{19}$$

Where represents the association between a branch and a node, with its direction flowing out of the node; and represents the association between a branch and a node, with its direction flowing into the node.

Taking phase A as an example, according to the law of energy conservation on the primary and secondary sides of the transformer, the voltage and current on the winding satisfy the following relationship:

$$\begin{cases} \dfrac{\dot{I}_{i-a}}{\dot{I}_{j-a}} = -\dfrac{1}{t_a} \\ \dfrac{\dot{U}_{i-a}}{\dot{U}_{j-a} - \dot{I}_{j-a}/y_a} = t_a \end{cases} \tag{20}$$

Here, $\dot{I}_{i-a}$ and $\dot{I}_{j-a}$ represent the branch currents flowing through the branches connected by $i-a$ and $j-a$, respectively. $y_a = 1/(R_a + \mathrm{j}X_a)$ represents the equivalent internal impedance of phase A of the transformer, t_a represents the turns ratio of phase A.

The relationship between branch current and node voltage is expressed as shown in Equation 21:

$$\begin{cases} \dot{I}_{i-a} = -\dfrac{1}{t_a}\dot{I}_{j-a} \\ \quad = -\dfrac{1}{t_a}\left(\dot{U}_{j-a} - \dfrac{\dot{U}_{i-a}}{t_a}\right)y_a \\ \quad = \dfrac{y_a}{t_a^2}\dot{U}_{i-a} - \dfrac{y_a}{t_a}\dot{U}_{j-a} \\ \dot{I}_{j-a} = \left(\dot{U}_{j-a} - \dfrac{\dot{U}_{i-a}}{t_a}\right)y_a \\ \quad = -\dfrac{y_a}{t_a}\dot{U}_{i-a} + y_a\dot{U}_{j-a} \end{cases} \tag{21}$$

Equation 20 is expressed as a matrix as shown in Equation 22:

$$\begin{bmatrix} \dot{I}_{i-a} \\ \dot{I}_{j-a} \end{bmatrix} = \begin{bmatrix} \dfrac{y_a}{t_a^2} & -\dfrac{y_a}{t_a} \\ -\dfrac{y_a}{t_a} & y_a \end{bmatrix} \begin{bmatrix} \dot{U}_{i-a} \\ \dot{U}_{j-a} \end{bmatrix} \tag{22}$$

Similarly, the relationship between the branch current and node voltage for phases B and C is derived, and the relationship between the three-phase branch current and voltage is ultimately obtained as follows:

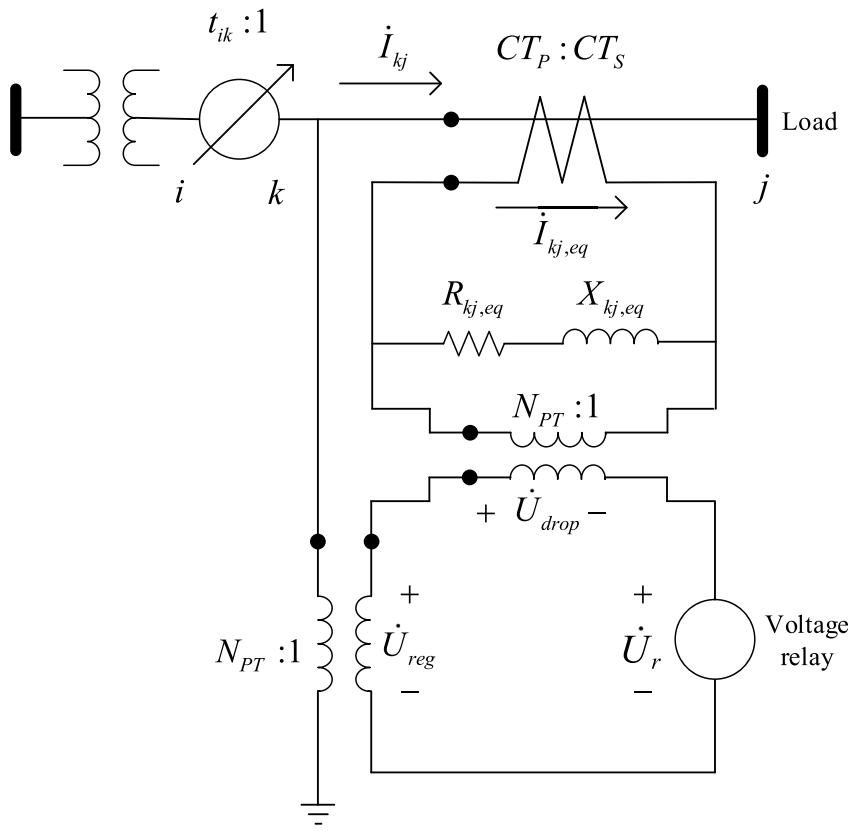


FIGURE 2
Detailed schematic diagram of OLTC voltage regulator part.

$$\begin{cases} \dot{I} = Y_{pr}\dot{U} \\ Y_{pr} = \begin{bmatrix} \frac{y_a}{t_a^2} & -\frac{y_a}{t_a} & & & & \\ -\frac{y_a}{t_a} & y_a & & & & \\ & & \frac{y_b}{t_b^2} & -\frac{y_b}{t_b} & & \\ & & -\frac{y_b}{t_b} & y_b & & \\ & & & & \frac{y_c}{t_c^2} & -\frac{y_c}{t_c} \\ & & & & -\frac{y_c}{t_c} & y_c \end{bmatrix} \end{cases} \tag{23}$$

In Equation 23: $\dot{I} = [\,\dot{I}_{i-a}\ \dot{I}_{j-a}\ \dot{I}_{i-b}\ \dot{I}_{j-b}\ \dot{I}_{i-c}\ \dot{I}_{j-c}\,]^T$ represents the branch current matrix of the OLTC, $\dot{U} = [\,\dot{U}_{i-a}\ \dot{U}_{j-a}\ \dot{U}_{i-b}\ \dot{U}_{j-b}\ \dot{U}_{i-c}\ \dot{U}_{j-c}\,]^T$ represents the node voltage matrix of the OLTC; Y_{pr} is defined as the original admittance matrix; t_b, t_c respectively represent the B and C phase transformation turns ratios.

By combining the original admittance matrix with the node-branch incidence matrix, the node admittance matrix is obtained from the following equation:

$$Y_{bus} = C^T Y_{pr} C$$

The branch power connected to each node is calculated using the node injection power expression as shown in Equations 24 and 25:

$$P_{i-\varphi}^{\mathrm{Reg}} = U_{i-\varphi} \sum_{\beta=\mathrm{A,B,C}} \left[U_{j-\beta} \left(G_{ij-\varphi\beta}^{\mathrm{Reg}} \cos\theta_{ij-\varphi\beta} + B_{ij-\varphi\beta}^{\mathrm{Reg}} \sin\theta_{ij-\varphi\beta} \right) \right] \tag{24}$$

$$Q_{i-\varphi}^{\mathrm{Reg}} = U_{i-\varphi} \sum_{j=\mathrm{A,B,C}} U_{j-\beta} \left(-B_{ij-\varphi\beta}^{\mathrm{Reg}} \cos\theta_{ij-\varphi\beta} + G_{ij-\varphi\beta}^{\mathrm{Reg}} \sin\theta_{ij-\varphi\beta} \right) \tag{25}$$

$G_{ij-\varphi\beta}^{\mathrm{Reg}}$ and $B_{ij-\varphi\beta}^{\mathrm{Reg}}$ represent the real and imaginary parts of the corresponding elements in the node admittance matrix Y_{bus} of the on-load tap changer, $G_{ij-\varphi\beta}^{\mathrm{Reg}}$ represents the mutual conductance between node $i-\varphi$ and node $j-\beta$, $B_{ij-\varphi\beta}^{\mathrm{Reg}}$ represents the mutual susceptance between node $i-\varphi$ and node $j-\beta$; The row name and column name of Y_{bus} are expressed as: node $i-a$, $i-b$, $i-c$, $j-a$, $j-b$, $j-c$.

The following describes the modelling of the regulator part in Figure 2.

The tap position of OLTC are regulated through a control circuit. When voltage control, direct control cannot be carried out on the high-voltage circuit. Therefore, voltage and current transformers are used to construct a simulated circuit - the control circuit. The control circuit is a proportional model of the actual line. For example, if the actual line transformer has a secondary side rated at 2.4 kV, the control circuit operates at a

voltage level of 120 V. By monitoring the voltage of the control circuit, the voltage at the load center of the actual line can be determined. If the voltage is below the normal level, indicating that the voltage at the loading center is below the position of the normal level, the tap changer is adjusted to raise the voltage at the loading center. For three-phase supply voltages below 10 kV, the allowable deviation is 7% of the rated voltage. For single-phase supply voltages of 220 V, the allowable deviation is +7% and −10% of the rated voltage. $R_{kj,eq}$ and $X_{kj,eq}$ are the proportional impedances of the actual line in the control circuit and are typically known quantities.

For the control circuit, the current rating of the primary winding of the current transformer is set as CT_P, and the current rating of the secondary winding is set as CT_S (usually taken as 5). The voltage transformer turns ratio is set as N_{PT}. First, according to Ohm's law, the equivalent impedance of the three-phase line is calculated as shown in Equation 26:

$$R_{line} + \mathrm{j}X_{line} = \frac{\dot{U}_k - \dot{U}_j}{\dot{I}_{kj}} \tag{26}$$

It should be noted that the equivalent impedance of the three-phase line is not the actual line impedance; it is the turns ratio of the actual measured voltage on the secondary side to the load center voltage difference to the load side current under the rated turns ratio.

The equivalent impedance of the line compensator is calculated from the equivalent impedance of the three-phase line as $R_{kj,eq}$、$X_{kj,eq}$ using the formula as shown in Equation 27:

$$R_{eq,kj} + \mathrm{j}X_{eq,kj} = \left(R_{line} + \mathrm{j}X_{line}\right) \cdot \frac{CT_P}{N_{PT} \cdot CT_S} \tag{27}$$

The current in the compensator branch is obtained from the actual line current $\dot{I}_{kj}$ and the current transformer $\dot{I}_{kj,eq}$.

$$\dot{I}_{kj,eq} = \dot{I}_{kj} \cdot \frac{CT_S}{CT_P} \tag{28}$$

In Equation 28: $\dot{I}_{kj} = [\ \dot{I}_{kj-a}\ \ \dot{I}_{kj-b}\ \ \dot{I}_{kj-c}\ ]^T$ represents the three-phase current in the line.

The voltage difference $\dot{U}_{drop}$ of the compensator branch is obtained as shown in Equation 29:

$$\dot{U}_{drop} = \left(R_{kj,eq} + \mathrm{j}X_{kj,eq}\right) \cdot \dot{I}_{kj,eq} \tag{29}$$

Finally, the control voltage of the compensator branch is as shown in Equation 30:

$$\dot{U}_r = \frac{\dot{U}_3}{N_{PT}} - \dot{U}_{drop} \tag{30}$$

Applying the droop control to the control voltage and tap changer turns ratio, taking phase A as an example, the principle of droop control for the OLTC is introduced. Define Δt_a, Δt_b, Δt_c as the adjustment amounts of the tap changer turns ratio for phases A, B, and C relative to their respective initial turns ratios, and add to the objective function as shown in Equation 31:

$$\min\ \ \left(\Delta t_a^2 + \Delta t_b^2 + \Delta t_c^2\right) \tag{31}$$

The physical meaning of the above equation is that when the control voltage is within the dead zone, the OLTC tap position remains unchanged.

The $\Delta t_a - U_{r-a}$ droop control function of the variation in the variable turns ratio variation and the control voltage is shown in Equation 31:

$$\Delta t_a = \begin{cases} \Delta t_{\max} - \Delta t_0 & U_{\min} \le U_{r-a} < U_l \\ k_{dr1}\left(U_{r-a} - U_{dbl}\right) & U_l \le U_{r-a} \le U_{dbl} \\ 0 & U_{dbl} < U_{r-a} < U_{dbh} \\ k_{dr2}\left(U_{r-a} - U_{dbh}\right) & U_{dbh} \le U_{r-a} \le U_h \\ t_{\min} - \Delta t_0 & U_h < U_{r-a} \le U_{\max} \end{cases} \tag{32}$$

The relationship between Δt_a and the control voltage U_{r-a} satisfies the following curve in Figure 3:

Where, $\Delta t_{\max}$ and $\Delta t_{\min}$ respectively represent the upper and lower limits of the variation in turns ratio Δt; Δt_0 represents the change in tap position corresponding to the initial turns ratio; U_{dbh} and U_{dbl} respectively represent the upper and lower bounds of the voltage dead zone; U_h and U_l respectively represent the upper and lower bounds of the voltage droop control curve; $U_{\max}$ and $U_{\min}$ respectively represent the upper and lower bounds of the bus voltage magnitude; k_{dr1} and k_{dr2} respectively represent the droop control coefficients, and $k_{dr1} < 0$, $k_{dr2} < 0$. When the voltage U_{r-a} is within the dead zone range, the OLTC does not actuate, and the variation in turns ratio is 0.

The relationship between the droop control coefficient and the variation in the turns ratio and the voltage is as shown in Equation 33:

$$\begin{cases} k_{dr1} = \dfrac{\Delta t_{\max}}{U_l - U_{dbl}} \\ k_{dr2} = \dfrac{-\Delta t_{\min}}{U_{dbh} - U_h} \end{cases} \tag{33}$$

From Figure 3, it can be seen that the characteristic curve represents a piecewise function. When the system operates near the turning point, the derivative is discontinuous, the algorithm search direction is uncertain, and it is difficult to smoothly switch operating curves, so the piecewise droop control function has strong non-smooth characteristics and is difficult to solve. Usually, mixed integer nonlinear programming is used to describe piecewise function control characteristics, but this method has low computational efficiency. In this paper, a fitting function is used to approximate the piecewise function, and the fitted droop control function is shown in Equation 34:

$$\begin{aligned} \Delta t_a = \Delta t_{\max} &+ \frac{k_{dr1}}{\alpha}\left[\ln\left(1 + e^{\alpha\left(U_{r-a} - U_l\right)}\right) - \ln\left(1 + e^{\alpha\left(U_{r-a} - U_{dbl}\right)}\right)\right] \\ &+ \frac{k_{dr2}}{\alpha}\left[\ln\left(1 + e^{\alpha\left(U_{r-a} - U_{dbh}\right)}\right) - \ln\left(1 + e^{\alpha\left(U_{r-a} - U_h\right)}\right)\right] \end{aligned} \tag{34}$$

Where, α is the fitting coefficient, set $\alpha = 500$ in this paper.

Approximately fitting the piecewise droop control function can transform it into a smooth function that is continuously differentiable, as shown in Figure 4. The smoothing function can avoid sudden changes in derivative order during algorithm iteration calculations, thus improving the convergence of the algorithm.

The variation of the ratio of the voltage regulator and the tap variable satisfies the relationship is shown in Equation 34:

$$\Delta t_a = x_{d-a} \cdot \Delta d \tag{35}$$

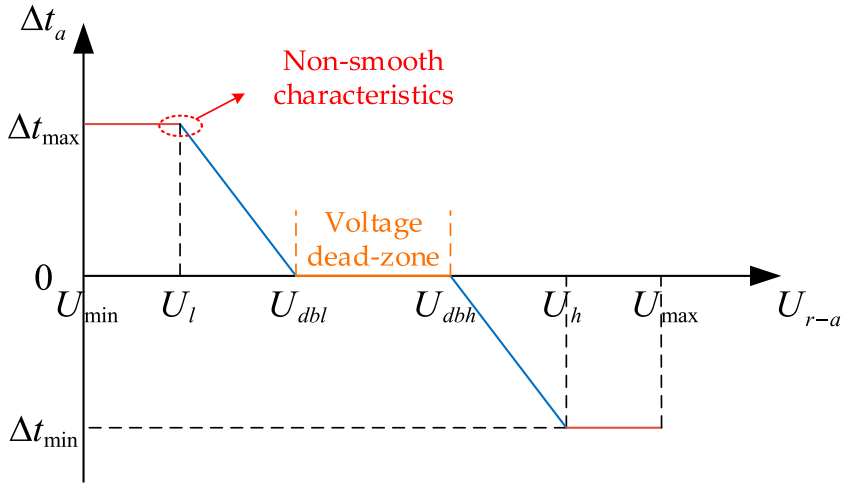


FIGURE 3
Voltage regulation control curve showing the variation of the turns ratio with control voltage droop.

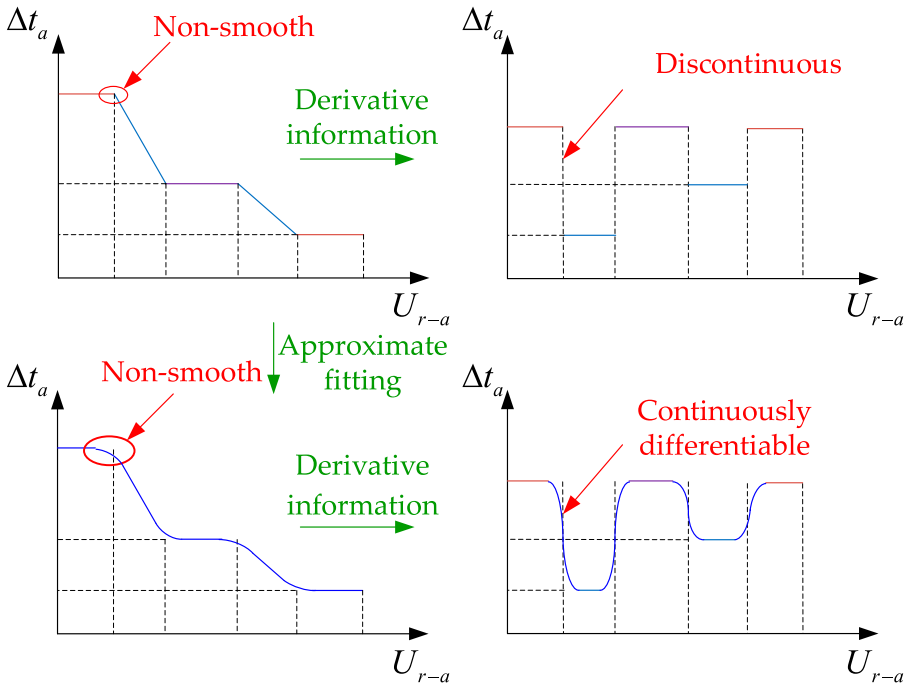


FIGURE 4
Derivative comparison diagram before and after fitting.

In the equation, x_{d-a} is the phase A tap-changer variable; Δd is the turns ratio change corresponding to a 1-tap adjustment, which is typically taken as $\Delta d = 0.00625$ p.u. in distribution networks.

In this paper, the discrete tap-changer variable is first treated as a continuous variable for state estimation calculations to obtain a continuous optimal solution for state estimation. Subsequently, a positive curvature quadratic penalty function is introduced in the objective function, and the state estimation is continued using the continuous optimal solution as the initial value to obtain an integer solution for the tap-changer setting.

Taking phase A as an example, the positive curvature quadratic penalty function is shown in Figure 5. x_{d0}、 x_{d1}、 x_{d2} are any three continuous discrete component values. Define the neighborhood $R(x_{d-a})$ as shown in Equation 36:

$$R(x_{d-a}) = \left\{ x_{d-a} \middle| x_{d1} - \frac{1}{2}\Delta d \le x_{d-a} \le x_{d1} + \frac{1}{2}\Delta d \right\} \tag{36}$$

Where, x_{d1} is the neighborhood center, which is the closest discrete grading value determined according to the continuous optimal solution.

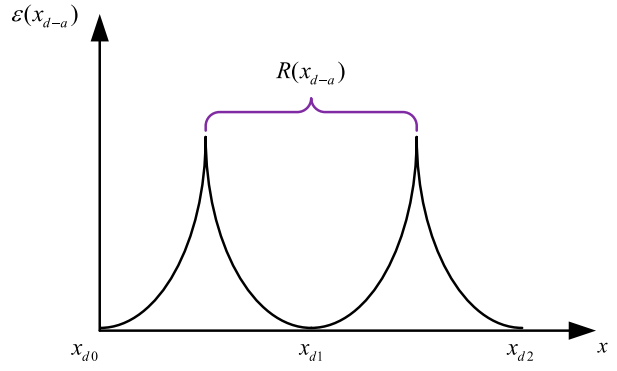


FIGURE 5
Quadratic penalty function model.

In the optimization process, when the value of x_{d-a} is in the neighborhood of the above definition, the penalty function $\varepsilon(x_{d-a})$ is introduced:

$$\varepsilon(x_{d-a}) = \frac{1}{2}\mu_d (x_{d-a} - x_{d1})^2 \tag{37}$$

In the equation, μ_d is the penalty factor, a known quantity; the penalty function will force x_{d-a} within the neighborhood to approach the neighborhood center. From Equation 37, we can obtain the first and second derivatives of the quadratic penalty function within the neighborhood.

$$\begin{cases} \dfrac{\partial \varepsilon(x_{d-a})}{\partial x_d} = \mu_d (x_{d-a} - x_{d1}) \\ \dfrac{\partial^2 \varepsilon(x_{d-a})}{\partial {x_d}^2} = \mu_d \end{cases} \tag{38}$$

From Equation 38, it can be observed that for the Newton method, the introduction of a penalty function in the objective function will result in the incremental inclusion of the first and second derivative terms in the Jacobian matrix and the Hessian matrix during the optimization iteration. Through this procedure, not only can rapid convergence of OLTC tap positions be achieved, but also the positive definiteness of the iteration matrix can be strengthened, thereby improving algorithm convergence. From a perspective in the field of power systems, this approach enables efficient adjustment of OLTC tap positions and enhances algorithm convergence by modifying derivative terms in iterative matrices.

3.2.2 Hybrid data-physical-driven linearization

Nonlinear constraints in the model of OLTC mainly fall into two categories: the first category is branch power constraints; the second category is droop control nonlinear constraints. Due to the introduction of the unknown variable t of the voltage regulator in the branch power constraint and the logarithmic function in the droop control nonlinear constraints, the aforementioned physical linearization methods are no longer applicable. In this paper, it is proposed to utilize a first-order Taylor expansion for physical linearization, followed by using Partial Least Squares Regression (PLSR) to calculate compensation errors, resulting in the development of a hybrid data-physical-driven Linearization model.

Firstly, introduce the principle of linearization of power constraints. For ease of description, define the relationship between the branch power of the transformer and the variables as shown in Equation 39:

$$\begin{cases} P_{i-\varphi}^{\mathrm{Reg}} = f_{i-\varphi}^{P}(X^{\mathrm{Reg}}) \\ Q_{i-\varphi}^{\mathrm{Reg}} = f_{i-\varphi}^{Q}(X^{\mathrm{Reg}}) \\ X^{\mathrm{Reg}} = \left[\, \theta_{i-\mathrm{ABC}} \;\; U_{i-\mathrm{ABC}} \;\; \theta_{j-\mathrm{ABC}} \;\; U_{j-\mathrm{ABC}} \;\; t_{\mathrm{ABC}} \,\right]^T \end{cases} \tag{39}$$

Where X^{Reg} represents the variables of the on-load tap-changing transformer, composed of the phase angle and voltage magnitude of the primary and secondary sides, and the three-phase turns ratio; $f_{i-\varphi}^{P}$ and $f_{i-\varphi}^{Q}$ respectively represent the functional relationships between the active and reactive power of phase φ branch and the variables of the on-load tap-changing transformer.

The first-order Taylor expansion is performed at the initial point $X^{\mathrm{Reg},0}$ of the variable, as shown in Equation 40:

$$\begin{cases} f_{i-\varphi}^{P}(X^{\mathrm{Reg}}) \approx f_{i-\varphi}^{P}(X^{\mathrm{Reg},0}) + f_{i-\varphi}^{P'}(X^{\mathrm{Reg},0}) \cdot (X^{\mathrm{Reg}} - X^{\mathrm{Reg},0}) + \Delta P_{i-\varphi}^{\mathrm{Reg}} \\ f_{i-\varphi}^{Q}(X^{\mathrm{Reg}}) \approx f_{i-\varphi}^{Q}(X^{\mathrm{Reg},0}) + f_{i-\varphi}^{Q'}(X^{\mathrm{Reg},0}) \cdot (X^{\mathrm{Reg}} - X^{\mathrm{Reg},0}) + \Delta Q_{i-\varphi}^{\mathrm{Reg}} \end{cases} \tag{40}$$

In the equations, $f_{i-\varphi}^{P'}$ and $f_{i-\varphi}^{Q'}$ represent the partial derivatives of the active and reactive power functions of the phase φ branch with respect to each variable in X^{Reg}; $\Delta P_{i-\varphi}^{\mathrm{Reg}}$ and $\Delta Q_{i-\varphi}^{\mathrm{Reg}}$ respectively represent the compensating errors of the active and reactive power of the phase φ branch, and the compensating errors are also calculated using the PLSR method.

Next, the linearization principle of the droop control function of the OLTC is introduced. Similar to the linearization principle of branch power, the functional relationship between the turns ratio change and the control voltage is defined in Equation 41:

$$\Delta t = f^{\Delta t}(U_r) \tag{41}$$

Where, $f^{\Delta t}$ represents the functional relationship between the variable ratio variation and the secondary side control voltage.

The first-order Taylor series expansion is performed at the initial value point $U_{r,0}$ of the variable, as shown in Equation 42:

$$f^{\Delta t}(U_r) \approx f^{\Delta t}(U_{r,0}) + f^{\Delta t'}(U_{r,0}) \cdot (U_r - U_{r,0}) + \Delta t_r \tag{42}$$

In the formula, $f^{\Delta t'}$ represents the derivative of the variable turns ratio variation function to the variable U_r ; Δt_r represents the compensation error of the variable turns ratio variation.

Then, the PLSR algorithm is used for data-driven error compensation, the specific principle and mathematical expressions are as follows:

The compensation error Δy of active power is defined in Equation 43:

$$\begin{cases} \Delta y = \xi y' + \eta \\ \Delta y = \left[\, \Delta P_{PV-\varphi}^{m} \;\; \Delta Q_{PV-\varphi}^{m} \,\right]^T \\ y' = \left[\, P_{load}^{PV-\varphi} \;\; Q_{load}^{PV-\varphi} \,\right]^T \end{cases} \tag{43}$$

Among them, ξ is the coefficient matrix and η is the constant matrix, which are obtained by PLSR fitting is the independent variable matrix of load composition; $P_{load}^{PV-\varphi}$ and $Q_{load}^{PV-\varphi}$ represent the vectors composed of active and reactive loads of the grid-connected node φ phase of the system.

The independent variable set and the dependent variable set can be obtained from the results of the power flow calculation. After obtaining the set of independent variables R and the dependent variable set Z, it is standardized.

$$\begin{cases} R^* = (R - \bar{R})S_R^{-1} \\ Z^* = (Z - \bar{Z})S_Z^{-1} \end{cases} \tag{44}$$

In Equation 44: R^*, Z^* represent the standardized independent variables and dependent variable sets; $\bar{R}$ and S_R respectively represent the mean and standard deviation of the independent variable set R.

For the standardized independent variable and dependent variable sets, the least squares regression analysis method is used to calculate the fitting coefficients, with the knowledge in the field of power systems:

$$C = \mathrm{PLS}(R^*, Z^*) \tag{45}$$

$$\eta = \bar{Z} - \frac{\bar{R}}{S_R} C \odot S_Z \tag{46}$$

In Equations 45 and 46: C represents the regression coefficient matrix obtained by PLSR. ξ and η represent the coefficient matrix and the constant matrix in the linear regression equation, respectively. Thus, the compensation error Δy can be obtained. Finally, the correction equation of the measurement is shown in Equation 47:

$$\begin{aligned} f_{PV-\varphi}^{m-P}(X_{PV}) &\approx f_{PV-\varphi}^{m-P}(X_{PV}^0) + \nabla f_{PV-\varphi}^{m-P}(X_{PV}^0) \times (X_{PV} - X_{PV}^0) + \Delta P_{PV-\varphi}^m \\ f_{PV-\varphi}^{m-Q}(X_{PV}) &\approx f_{PV-\varphi}^{m-Q}(X_{PV}^0) + f_{PV-\varphi}^{m-Q}(X_{PV}^0) \times (X_{PV} - X_{PV}^0) + \Delta Q_{PV-\varphi}^m \end{aligned} \tag{47}$$

3.3 Bad data identification model based on MILP

For the suspicious measurement set, the MILP bad data identification model is constructed based on the hyhrid data-physical driven linearization method as shown in Equations 48–51:

$$\min \quad \sum_{i=1}^{m} b_i \tag{48}$$

$$s.t. \quad c^L(x) = 0 \tag{49}$$

$$h_i^L(x) - z_i^{\mathrm{m}} \geq -3\sigma_i - \mathrm{M}b_i \tag{50}$$

$$h_i^L(x) - z_i^{\mathrm{m}} \leq 3\sigma_i + \mathrm{M}b_i \tag{51}$$

Where, $c^L(x)$ represents the linearized equality constraint, and $h_i^L(x)$ represents the linearized measurement equation of the line.

When z_i is a normal measurement, that is, it does not belong to the suspicious measurement set, Formulas 50, 51 in the model should be rewritten as shown in Equations 52 and 53:

$$h_i^L(x) - z_i^{\mathrm{m}} \geq -3\sigma_i \tag{52}$$

$$h_i^L(x) - z_i^{\mathrm{m}} \leq 3\sigma_i \tag{53}$$

The MILP bad data identification method is not affected by the leverage point, and can quickly and accurately identify the bad data in the leverage measurement.

4 Case study

Due to the need for further improvement in the accuracy and computational efficiency of the current bad data identification method in distribution networks, this paper proposes a bad data identification method based on MILP model to simultaneously address the issues of model accuracy and computational efficiency. In this chapter, a typical low-voltage 42-node distribution network case study is used to validate the effectiveness of the proposed bad data identification model. The relevant case studies are conducted on the MATLAB software platform, utilizing an Intel(R) Core(TM) i5-10210U CPU 1.60 GHz processor. The simulation calculations are carried out using per unit values, with a base voltage of 13.8 kV and a base power of 100 kVA for the case study system. The state estimation calculations are initialized in a flat start manner, with measurement values subject to Gaussian distribution errors with mean of 0 and variance of σ^2 added to the true values. The standard deviations for voltage measurements, branch power measurements, and node injection power measurements are set as $\sigma_i = 0.004$, $\sigma_{\mathrm{Bran}} = 0.008$, and $\sigma_{\mathrm{Node}} = 0.01$.

For ease of analysis, this paper selects two mathematical indicators, the Root Mean Square Error (RMSE) and the Maximum Absolute Error (MAE). In this paper, RMSE represents the square root of the ratio of the sum of the squares of errors between the estimated values and the true values to the data dimension; MAE is generally used to measure the range of absolute errors, i.e., the maximum absolute error between the estimated values and the true values. The mathematical expressions for these two indicators are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i}^{n}(x_i - \hat{x}_i)^2} \tag{54}$$

$$\mathrm{MAE} = \max\{|x_i - \hat{x}_i|\} \quad i = 1, 2, 3 \ldots \tag{55}$$

In the equation, $\hat{x}_i$ represents the true value of the state variable x_i.

To verify the model, the modified IEEE-33 node test feeder is used to test the linearization and bad data identification model proposed in this paper.

The schematic diagram of the modified IEEE-33 node test feeder is shown in Figure 6. The feeder consists of 33 nodes and 32 branches. The power flow results of the system are used as the normal measurements for the entire system, including 402 measurements, including 6 voltage magnitude measurements, 198 branch power measurements, and 198 node injection power measurements.

4.1 Linearization accuracy comparison

Firstly, test the hybrid data-physical-driven linearization model proposed in this paper. Based on the IEEE 33-node parameters, training and test datasets can be generated from the AC power flow model. The case study sets training and testing data as simulated data randomly generated within the range of 95%–105% of the actual load after removing outliers, with a total of 100 sets of training samples. The measurement linearization accuracy obtained from

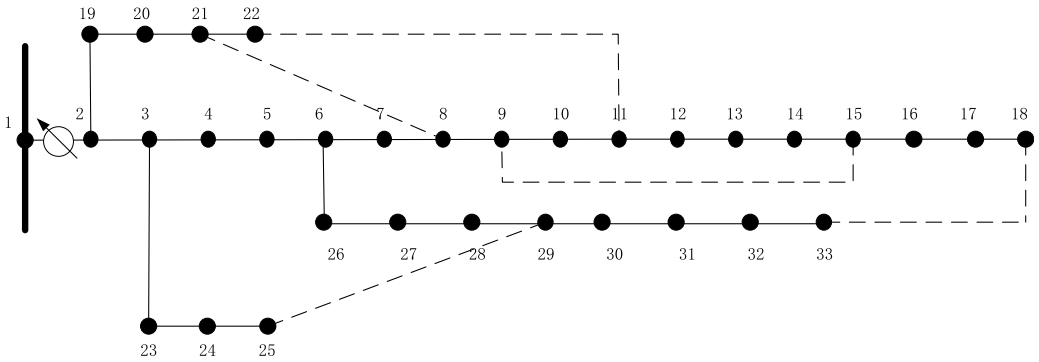


FIGURE 6
A modified IEEE-33 node test feeder.

TABLE 1 Comparison of the accuracy between linearization methods.

Linearization method		RMSE	MAE
Hybrid data-physical-driven	LSR	1.36×10^{-2}	0.6104
	PLSR	4.326×10^{-11}	3.593×10^{-5}
	BLR	1.4784×10^{-6}	1.16×10^{-2}
Physical driven linearization	—	7.2×10^{-3}	0.410

three hybrid data-physical-driven methods, namely, Taylor expansion linearization, PLSR regression, Least squares regression (LSR) (Shao et al., 2023), and Bayesian linear regression (BLR) (Liu et al., 2019), are compared. To facilitate analysis, the difference between the linearized power flow results and the true values is represented by two mathematical indicators, RMSE and MAE, as shown in Equations 54, 55 respectively. The comparison results in Table 1.

As shown in Table 1, the hybrid data-physical-driven Linearization method proposed in this paper has a higher advantage in terms of linearization accuracy. Compared to physical-driven linearization, This method can improve the accuracy by 10^8 of magnitude. Additionally, compared to data-driven error compensation methods such as LSR and BLR, the PLSR method used in this paper has the highest accuracy and is closer to the results of nonlinear constraints.

4.2 OLTC tap position analysis

The OLTC tap position is verified by introducing a positive curvature quadratic penalty function into the tap position alignment model proposed in this paper. The tap position is treated as a continuous variable for state estimation calculations, obtaining a continuous optimal state estimate. The positive curvature quadratic penalty function is introduced into the objective function, using the continuous optimal solution as the initial value to continue the state estimation and obtain an integer solution for the tap position. Two different tap position processing models for on-load tap changing are set up to verify the effectiveness of the proposed model.

Model 1: Traditional direct treatment of the tap position as a continuous variable.

Model 2: Tap position rounding method based on positive curvature quadratic penalty function for OLTC.

The accuracy of voltage magnitude estimation for the two models is shown in Table 2.

The results indicate that for the OLTC model, the method of rounding tap positions with a positive curvature quadratic penalty function can lead to state estimation results with higher accuracy.

4.3 Analysis of bad data identification results

4.3.1 Comparative analysis of traditional statistical methods for bad data identification

According to the two-stage bad data identification model proposed in this paper, 10 bad data are set as shown in Table 3. At the same time, the false alarm rate is set to $P_e = 0.0025$. By querying the standard normal distribution table, the normal range of normalized residuals can be obtained, and the

TABLE 2 Voltage magnitude estimation results under different OLTC tap position handling methods.

Model	RMSE			MAE		
	Phase A	Phase B	Phase C	Phase A	Phase B	Phase C
Model 1	1.119×10^{-2}	3.063×10^{-3}	8.643×10^{-3}	0.125	3.238×10^{-3}	0.140
Model 2	5.405×10^{-3}	4.968×10^{-4}	2.944×10^{-3}	0.09	3.469×10^{-2}	0.094

TABLE 3 Detail information of the bad data and the corresponding projection statistics and normalized residual values.

Measurement number	Measurement name	True value (p.u.)	Bad data (p.u.)	D_i	r_i
19	branch powerP2-3-a	4.693	8.9	0.422	5.367
25	branch powerP3-4-b	4.553	10	0.936	7.219
33	branch powerP3-5-b	4.006	9.65	0.690	7.704
38	branch powerP6-7-b	3.539	7.37	0.484	5.113
45	branch powerQ7-8-a	2.854	5.49	1.201	2.222
59	branch powerQ9-11-c	0.919	3.377	1.711	2.408
66	branch powerQ19-20-c	0.844	8.3	0.334	9.359
93	branch powerQ22-27-b	0.863	7.68	0.344	6.936
164	branch powerP36-38-b	0.436	5.5	0.970	9.140
172	branch powerQ39-40-c	0.045	6.7	0.251	8.354

TABLE 4 Comparison of results of different identification methods.

Identification method	Stage 1		Stage 2
	Projection statistics method	Maximum normalized residual method	MILP
Identification result (pcs)	265	8	10

TABLE 5 Comparison of test results of different bad data identification methods.

Identification method	Identification model	Solver	Identification Result	CPU time-consuming/s
Two-Stage	Test 1	CPLEX (Kia et al., 2016)	True	0.546
		BARON (Ghildyal and Sahinidis, 2001)	True	0.702
		OSICPLEX (Apland and Sun, 2019)	True	0.159
		OSIMOSEK (Baradar and Hesamzadeh, 2014)	True	1.234
		SCIP (Vigerske and Gleixner, 2017)	True	1.716
	Test 2	BONMIN (Gupta and Ravindran, 1985)	True	1,000.00(Over the maximum time limit)
Single-Stage	Test 3	CPLEX	True	0.826
		BARON	True	1.362
		OSICPLEX	True	1.011
		OSIMOSEK	True	1.911
		SCIP	True	39.700
	Test 4	BONMIN	True	1,000.00(Over the maximum time limit)

detection threshold τ is set to 3 and the threshold D for the projection statistics is set to 1.

According to Table 4, among the 10 bad data, the projection statistical method effectively identified only 8 pcs bad data, while all other measurements with projection statistics values greater than 1 resulted in false alarms. Similarly, the maximum normalized residual method also failed to accurately identify the bad data. Therefore, both traditional statistical methods for bad data identification have certain limitations.

4.3.2 Comparative analysis of single-stage and two-stage bad data identification methods

Due to the potential for misjudgments and missed detections by traditional statistical methods when multiple measurement bad data points are present in the system, these methods may not accurately identify measurement bad data. In contrast, the MILP bad data identification model used in the second stage can accurately identify all measurement bad data in the set identified by traditional methods when misjudgments or missed detections occur in the first stage. Therefore, in the first stage of this paper, all 265 lever measurements in the system were placed into a suspicious measurement set. The maximum normalized residual method was then used to identify the remaining 137 measurements, thereby validating the accuracy of the proposed bad data identification model and its efficiency compared to the single-stage model. Four sets of tests were conducted to compare the bad data identification method proposed in this paper with the traditional MINLP-based bad data identification method:

Test 1: Compression of the suspicious measurement set using the MILP-based two-stage bad data identification method proposed in this paper;

Test 2: Compression of the suspicious measurement set using the traditional MINLP-based two-stage bad data identification method;

Test 3: No compression of the suspicious measurement set, directly using the MILP-based single-stage bad data identification method;

Test 4: No compression of the suspicious measurement set, using the traditional MINLP-based single-stage bad data identification method.

Models were built for both methods in the GAMS optimization software, different solvers were called for calculation, and the results were compared and analyzed.

According to Table 5, for the modified IEEE-33 node test feeder case chosen in this paper, the two-stage model reduces the number of 0/1 integer variables and improves identification accuracy. However, for traditional MINLP-based bad data identification models, feasible solutions could not be obtained with most solvers, with correct results only achievable using the BONMIN solver. In contrast, the MILP-based bad data identification model achieved accurate identification results with most solvers. In terms of identification efficiency, even with the two-stage model, the MINLP identification model's solving time reached the computational limit of 1,000 s, and the solver exited abnormally, indicating limited applicability. For the MILP-based bad data identification method, using the two-stage model, the SCIP solver improved solving efficiency by approximately 23 times, and the CPLEX solver achieved the highest efficiency, requiring only 0.546 s.

5 Conclusion

To cope with the issue that existing traditional identification methods do not consider the nonlinear and discrete characteristics of on-load tap changers, making it difficult to achieve accurate and efficient identification of bad data., This paper proposes a MILP-based two stage bad data identification method. Detailed control characteristics of on-load tap changers are modelled, and a positive curvature quadratic penalty function is introduced to achieve fast tap normalization. In the first stage, leveraging projection statistics and maximum normalization residue methods effectively identifies leverage points and suspicious bad data, reduces the set of suspicious measurements. In the second stage, by linearizing nonlinear constraints and solving bad data identification model based on the MILP, the efficiency of the solution is greatly enhanced. The proposed model can achieve efficient and accurate identification of bad data, while ensuring optimal solutions by introducing penalty functions into the objective function for effective tap normalization.

Data availability statement

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

Author contributions

SH: Conceptualization, Funding acquisition, Methodology, Writing–original draft. CR: Formal Analysis, Investigation, Writing–original draft. MZ: Project administration, Resources, Writing–review and editing. LC: Software, Supervision, Writing–review and editing. YM: Validation, Visualization, Writing–review and editing. TZ: Validation, Visualization, 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 research was funded by State Grid Hebei Electric Power Co., Ltd. science and technology projects, grant number kj2023-076.

Conflict of interest

Authors SH, CR, and LC were employed by Economic and Technology Research Institute of State Grid Hebei 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.

The authors declare that this study received funding from State Grid Hebei Electric Power Co.,Ltd. The funder had the following involvement in the study: the study design, collection, analysis, interpretation of data, and the writing of this article.

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

Apland, J., and Sun, B. (2019). A Multi-Period, multiple objective, mixed integer programming, GAMS model for transit system planning

Baradar, M., and Hesamzadeh, M. R. (2014). A stochastic SOCP optimal power flow with wind power uncertainty. 1–5. doi:10.1109/PESGM.2014.6939790

Chen, Y. (2021). *Power system state estimation*. Beijing, China: Science Press.

Ghildyal, V., and Sahinidis, N. V. (2001). "Solving global optimization problems with baron," in *From Local to Global Optimization. Nonconvex Optimization and Its Applications*. Editors A. Migdalas, P. M. Pardalos, and P. Värbrand (Boston, MA: Springer. doi:10.1007/978-1-4757-5284-7_10

Gupta, O. K., and Ravindran, A. (1985). Branch and bound experiments in convex nonlinear integer programming. *Manage. Sci.* 31 (12), 1533–1546. doi:10.1287/mnsc.31.12.1533

Handschin, E., and Kliokys, E. (1995). Transformer tap position estimation and bad data detection using dynamic signal modelling. *IEEE T. Power Syst.* 10 (2), 810–817. doi:10.1109/59.387921

Handschin, E., Schweppe, F. C., Kohlas, J., and Fiechter, A. (1975). Bad data analysis for power system state estimation. *IEEE Trans. Power Apparatus Syst.* 94 (2), 329–337. doi:10.1109/T-PAS.1975.31858

Huang, S. J., and Lin, J. M. (2004). Enhancement of anomalous data mining in power system predicting-aided state estimation. *IEEE T. Power Syst.* 19 (1), 610–619. doi:10.1109/TPWRS.2003.818726

Irving, M. R. (2008). Robust state estimation using mixed integer programming. *IEEE T. Power Syst.* 23 (3), 1519–1520. doi:10.1109/TPWRS.2008.926721

Ju, Y., and Huang, Y. (2023). State estimation for an AC/DC hybrid power system adapted to non-smooth characteristics. *Power Syst. Prot. Control* 51, 141–150. doi:10.19783/j.cnki.pspc.220368

Kia, R., Shahnazari-Shahrezaei, P., and Zabihi, S. (2016). Solving a multi-objective mathematical model for a multi-skilled project scheduling problem by CPLEX solver. *IEEE*, 1220–1224. doi:10.1109/IEEM.2016.7798072

Korres, G. N., Katsikas, P. J., and Contaxis, G. C. (2004). Transformer tap setting observability in state estimation. *IEEE T. Power Syst.* 19 (2), 699–706. doi:10.1109/TPWRS.2003.821629

Lin, Y., and Abur, A. (2018). A highly efficient bad data identification approach for very large scale power systems. *IEEE T. Power Syst.* 33 (6), 5979–5989. doi:10.1109/TPWRS.2018.2826980

Liu, Y., Zhang, N., Wang, Y., Yang, J., and Kang, C. (2019). Data-Driven power flow linearization: a regression approach. *IEEE T. Smart Grid* 10 (3), 2569–2580. doi:10.1109/TSG.2018.2805169

Maalouf, H., Jabr, R. A., and Awad, M. (2013). Mixed-integer quadratic programming based rounding technique for power system state estimation with discrete and continuous variables. *Electr. Pow. Compo. Sys.* 41 (5), 555–567. doi:10.1080/15325008.2012.755233

Mili, L., Cheniae, M. G., Vichare, N. S., and Rousseeuw, P. J. (1996). Robust state estimation based on projection statistics [of power systems]. *IEEE T. Power Syst.* 11, 1118–1127. doi:10.1109/59.496203

Nanchian, S., Majumdar, A., and Pal, B. (2017). Three-Phase state estimation using hybrid particle swarm optimization. *IEEE T. Smart Grid.* 8 (3), 1035–1045. doi:10.1109/TSG.2015.2428172

Shao, Z., Zhai, Q., and Guan, X. (2023). Physical-Model-Aided Data-Driven linear power flow model: an approach to address missing training data. *IEEE T. Power Syst.* 38 (3), 2970–2973. doi:10.1109/TPWRS.2023.3256120

Shiroi, M., and Hosseinie, S. H. (2008). "Observability and estimation of transformer tap setting with minimal PMU placement," in *IEEE power and energy society general meeting - conversion and delivery of electrical energy in the 21st century*. Paper presented at the 2008. doi:10.1109/PES.2008.4596382

Teixeira, P. A., Brammer, S. R., Rutz, W. L., Merritt, W. C., and Salmonsen, J. L. (1992). State estimation of voltage and phase-shift transformer tap settings. *IEEE T. Power Syst.* 7 (3), 1386–1393. doi:10.1109/59.207358

Vigerske, S., and Gleixner, A. (2017). SCIP: global optimization of mixed-integer nonlinear programs in a branch-and-cut framework. *Optim. Methods Softw.* 33 (3), 563–593. doi:10.1080/10556788.2017.1335312

Zhao, J., and Mili, L. (2018). Vulnerability of the largest normalized residual statistical test to leverage points. *IEEE T. Power Syst.* 33 (4), 4643–4646. doi:10.1109/TPWRS.2018.2831453

Zhuang, F., and Balasubramanian, R. (1987). Bad data processing in power system state estimation by direct data deletion and hypothesis tests. *IEEE Power Eng. Rev.* PER-7 (5), 40. doi:10.1109/MPER.1987.5527244

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 16 October 2024
DOI 10.3389/fenrg.2024.1474969

Check for updates

OPEN ACCESS

EDITED BY
Yang Yu,
Nanjing University of Posts and Telecommunications, China

REVIEWED BY
Minh Quan Duong,
The University of Danang, Vietnam
Lefeng Cheng,
Guangzhou University, China

*CORRESPONDENCE
Chunxiang Yang,
2071763385@qq.com

RECEIVED 02 August 2024
ACCEPTED 25 September 2024
PUBLISHED 16 October 2024

CITATION
Yang C, Wu G, Zhang Y, Bao G and Wang J (2024) A study of short-term wind power segmentation forecasting method considering weather on ramp segments.
Front. Energy Res. 12:1474969.
doi: 10.3389/fenrg.2024.1474969

COPYRIGHT
© 2024 Yang, Wu, Zhang, Bao and Wang. 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 study of short-term wind power segmentation forecasting method considering weather on ramp segments

Chunxiang Yang[1]*, Guodong Wu[1,2], Yongrui Zhang[3], Guangqing Bao[4] and Jianhui Wang[5]

[1]Power Dispatch Center of State Grid Gansu Electric Power Company, Lanzhou, China, [2]Department of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou, China, [3]Electric Power Science Research Institute of State Grid Gansu Electric Power Company, Lanzhou, China, [4]School of Electronics and Information Engineering Southwest Petroleum University, Chengdu, China, [5]School of Electrical Engineering, Northwest University for Nationalities, Lanzhou, China

The short-term fluctuation of wind power can affect its prediction accuracy. Thus, a short-term segmentation prediction method of wind power based on ramp segment division is proposed. A time-series trend extraction method based on moving average iteration is proposed on the full-time period to analyze the real-time change characteristics of power time-series initially; secondly, a ramp segment extraction method based on its definition and identification technique is proposed based on the results of the trend extraction; and a segmentation prediction scheme is proposed to lean the power prediction under different time-series: the LightGBM-LSTM is proposed for the non-ramping segment using point prediction, and the CNN-BiGRU-KDE is proposed for probabilistic prediction of ramp segments. From the results, this ramp segment definition and identification technique can effectively identify the ramp process of wind power, which makes up for the misidentification and omission of the classical climbing event definition; meanwhile, the segment prediction scheme not only meets the prediction accuracy requirements of the non-ramping segment, but also provides the effective robust information for the prediction of the ramping period, which offers reliable reference information for the actual wind farms. In particular, it is well adapted to wind power prediction under extreme working conditions caused by ramping weather, which is a useful addition to short-term wind power prediction research.

KEYWORDS

ramp segment, wind power, trend identification, probabilistic fitting, segmental prediction

1 Introduction

In 2020, with the deepening understanding of the "dual-carbon" goal by all parties in society, China has put forward the goal of "2030 carbon peak, 2060 carbon neutral" (State Grid, 2021), and wind power is ushering in rapid development. Wind power is difficult to predict due to its unique stochasticity and instability, which poses a great challenge to the reliable operation of wind farms and smart energy systems. With the development of China's power marketization, accurate and efficient short-term wind power prediction is especially important to enhance the capacity of wind power consumption and promote the

efficient interaction of source, grid, and load (Lefeng et al., 2022; Lefeng et al., 2021; Lefeng et al., 2020). Currently, there is a lack of in-depth research on the short-term power prediction method for wind power over the whole period considering the weather in the climbing segment (Cheng and Yu, 2019).

Existing short-term power prediction methods for wind power can be divided into two kinds: the physical method (Ernst et al., 2007) and the data-driven method (Wang et al., 2022). Among them, the former is built based on the atmospheric motion portrayal, according to the meteorological environment, geographic factors, and other information, the use of hydrodynamics and other physical laws to establish a model, focusing on the optimization of the boundary conditions and physical solution rules, with the characteristics of modeling difficulties, large computational volume, and therefore poor timeliness, is generally suitable for medium- and long-term forecasting (Cassola and Burlando, 2012). The latter takes the establishment of linear or nonlinear mapping between relevant meteorological features and power time series as the main means, emphasizes the search for intrinsic laws from multi-source, multi-dimensional, and multi-modal data, and has been widely used because of its better prediction accuracy (Zhou et al., 2021).

Existing studies usually categorize data-driven methods into deterministic forecasting and uncertainty forecasting based on the result presentation. Among them, the existing deterministic wind power prediction methods mainly include Auto-regressive and moving average (ARMA) (Erdem and Shi, 2011), Convolutional Neural Networks (CNN) (Men et al., 2016), and vector machines (Hu et al., 2014). Deterministic methods can form a mathematical abstract mapping relationship between inputs and outputs through data mining and machine learning and are suitable for power time series with gentle curve fluctuations. Specifically, literature (Meng et al., 2021) proposes a parameter optimization-based attention mechanism for accelerating the early prediction model to mine the temporal correlation of the input series-gated recurrent unit (GRU) short-term wind power prediction model; literature (Zhou et al., 2021) proposes a wind power prediction model that introduces the volatility hierarchical error correction model, which is based on the improvement of long-term recurrent convolutional neural network. All of the above literature has improved the prediction accuracy to a certain extent, and better prediction results can be achieved under normal fluctuating power hours. However, in the face of the ramping section of the weather under the fluctuating power, a single use of the above deterministic prediction methods will not be able to quantify the prediction error, and the stability of the prediction results is poor, and the combination of prediction techniques is applied and born (Gao et al., 2016; Liu et al., 2024). At the same time, due to the more complex and variable wind power scenarios, it poses a more serious challenge to the prediction methods. For this reason, we have carried out an in-depth study of the problems and difficulties existing in the current wind power prediction work.

Uncertainty prediction is a probabilistic interval prediction method represented by kernel density estimation (KDE) (Wang et al., 2024; Haoyi et al., 2023). Uncertainty prediction considers the randomness of the results, quantifies the prediction error, provides more information compared to the traditional point estimation, and can significantly improve the effectiveness of power hour prediction under weather in the climbing section (Jianhou et al., 2024). Specifically, literature (Wang et al., 2024) introduces a new offshore wind speed point and interval prediction model that combines an innovative two-layer decomposition technique, GRU and KDE. However, the lack of a typical power scenario delineation leads to a low prediction accuracy of the model for some power periods. Literature (Zareipour et al., 2011a; Cui et al., 2019) proposes data-driven probabilistic wind power ramp prediction methods based on massive simulated scenarios, but such models have yet to improve their robustness under weather in the ramp section. Literature (Ouyang et al., 2019) proposes an integrated learning method to generate probabilistic prediction results, but the method does not take into account the interference of power timing pseudo-inflection points on the complete extraction of the ramp segment period and does not highlight the improvement of the model's accuracy under ramp segment weather. The above uncertainty prediction method improves the performance of wind power prediction under complex meteorological conditions to a certain extent, but there are still the following shortcomings: first, the lack of targeted optimization of the ramp segment of the extreme weather caused by the sudden change of power scenarios, which affects the prediction accuracy; second, the deterministic prediction method of the gentle power period is sufficient to meet the demand for prediction accuracy and stability, and the uncertain prediction takes up a large number of computing resources and the prediction interval under the gentle power period is too long to meet the prediction accuracy and stability requirements. Second, the deterministic prediction method is sufficient to meet the demand for prediction accuracy and stability in the gentle power period, while the uncertainty prediction takes up a lot of computing resources and the prediction interval is too large in the gentle power period, which affects the reasonableness and intuition of prediction.

The basis of ramp prediction is its identification technology. There have been in-depth studies on the research of wind power ramp events abroad, but the definition of it by various research institutions has not yet formed a unified standard. Literature (Potter et al., 2009; Ferreira et al., 2011) summarized four different definitions of ramp events by considering several factors such as power amplitude change, duration, and ramping rate. According to Truewind (2008), the occurrence of a "ramp event" is accompanied by a large change in wind speed in a short period, and the larger the amplitude change, the smaller the duration, and the faster the ramping rate, the more serious the ramp event is. Common studies set the minimum threshold of climb duration at 1 h, but ramp events of less than 1 h are also possible (Kamath, 2010; Kamath, 2011). Further, the literature (Zheng and Kusiak, 2009; Zareipour et al., 2011b) used a mean clustering algorithm and support vector machine to classify the ramp events in the historical data, respectively, and analyzed the characteristics and hazards of different types of ramp events.

The recognition technology of ramp events in China is not mature, and it is based on power prediction. Literature (Greaves et al., 2009) used a numerical weather prediction system to identify possible future wind power's ramp events by obtaining meteorological background information. In literature (Cui et al., 2014; Huang et al., 2016; Ouyang et al., 2017), ARMA, Kalman, and neural network models were used to predict the power first, and then

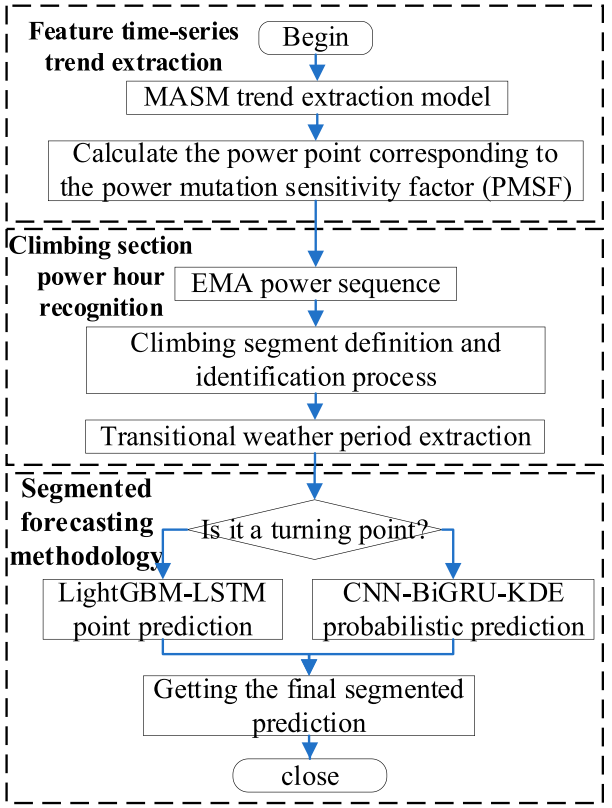


FIGURE 1
Method flow chart.

the predicted power was used for ramp recognition. Due to the lack of consideration of the characteristics of ramp events, the effectiveness of these methods in the identification of ramp events is very limited, which makes the identification of ramp events one of the urgent problems to be solved in the grid connection of wind power.

Aiming at the above deficiencies, a short-term wind power segmentation prediction method based on ramp period division is proposed in this paper. Specifically, a trend extraction model based on the moving average sensitivity method (MASM) is first proposed for the whole period to characterize the real-time change of power time series preliminarily; furthermore, a hill-ramp segment definition and identification method is proposed to extract the hill-ramp power period for the sub-time period; finally, a segmented prediction method is proposed to make lean prediction of wind power for the whole period: a light gradient boosting machine (LightGBM) - Long short-term memory (LSTM) is proposed for the non-hill-ramp segment period. Finally, a segmented prediction method is proposed to make a lean prediction of wind power for the whole period: a LightGBM-LSTM combination prediction method is proposed for the non-ramp period; a probabilistic prediction method based on CNN-BiGRU-KDE is proposed for the ramp period. The experimental results show that the prediction accuracy of the method proposed in this paper is greatly improved compared with the existing methods, providing new ideas for short-term wind power prediction.

2 Basic idea

Changes in meteorological parameters under ramp segment weather are characterized by instantaneous sudden changes and drastic amplitude, which leads to many problems in wind power prediction under ramp segment weather conditions.

The main problems are as follows:

(1) As the weather in the ramp segment has various changes in meteorological patterns in a short period, it is easy to cause misjudgment of the trend, which affects the accuracy of power extraction in the ramp segment.
(2) Due to the strong stochasticity and complexity of the weather mutation period in the ramp segment, it is difficult to accurately and completely extract the power mutation period in the ramp segment by the power mutation period extraction method with the fixed characteristics as the extraction factor.
(3) Different meteorological models correspond to different time series characteristics in the ramp weather period. To fully utilize the performance advantages of deterministic and uncertainty prediction methods, it is one of the urgent problems to propose a segmented prediction strategy to match different weather patterns.

To address the above issues specifically, this paper proposes a segmented prediction method based on ramp segment identification and recognition technology. The specific method flow is shown in Figure 1. Firstly, the MASM model is used to extract the trend components in the time series, smoothing transitions and capturing trend changes. Subsequently, the power mutation sensitivity factor (PMSF) is employed to calculate specific power points of sudden changes, identify key changes in the time series, and obtain the exponential moving average (EMA) sequence. On this basis, ramp segments in EMA are identified and extracted by defining ramp segments and setting ramp thresholds (ramp amplitude, ramp rate), i.e., periods of significant changes in weather conditions, to analyze their impact on electricity demand. In the prediction phase, a segmented prediction method is adopted, processing the time series according to different characteristics or patterns.

For non-ramp segments, the LightGBM-LSTM model is used for point prediction, combining the advantages of gradient boosting and long short-term memory networks to capture complex patterns and temporal dependencies. For ramp segments, the CNN-BiGRU-KDE model is employed for probabilistic prediction, generating predictive probability distributions. Finally, the results of point prediction and probabilistic prediction are integrated to form the final segmented prediction. This method not only considers the accuracy of prediction but also incorporates prediction uncertainty, providing more comprehensive and reliable prediction results.

3 Time-series trend extraction

Under the weather of the ramp segment, the wind power shows drastic changes, to accurately identify the ramp power period, the first step is to extract the time series trend of the whole period. The traditional method of recognizing the time series mutation is to extract the index parameters such as mutation amplitude and

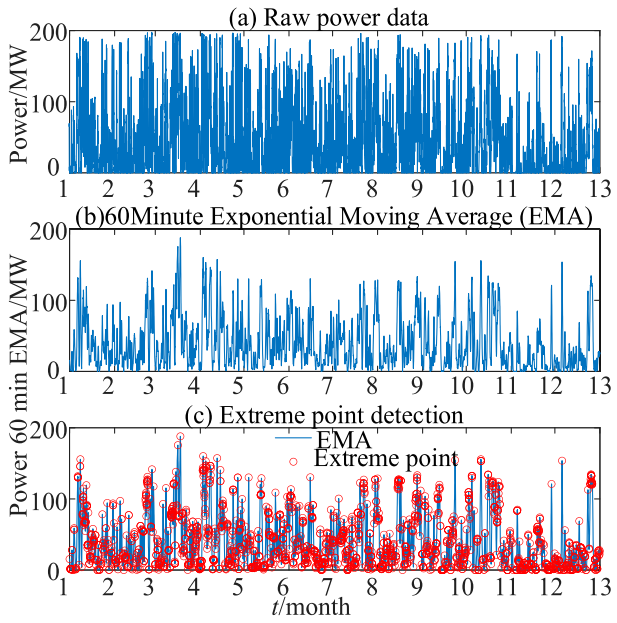


FIGURE 2
Schematic diagram of the effect of EMA and PMSF. **(A)** Raw power data. **(B)** 60 min Expontial Movieng Average (EMA). **(C)** Extreme Point detection.

mutation duration as the basis of identification. This method is only applicable to a single time-sequence mutation scenario, and it is easy to cause insufficient extraction of the mutation period for the complex and variable time-sequence mutation scenarios of the ramp segment. To establish an ideal early warning mechanism for weather periods in the ramp segment, this paper proposes a novel method of describing the time-sequence trend by taking the historical time-sequence characteristics into full consideration. Different from the traditional trend extraction method that directly takes the original time series as the feature extraction object, this method takes the moving average as the trend research object. It not only avoids the trend misjudgment caused by the raw power time series noise but also retains the timeliness of the time series change trend. This method extracts the time-series trend from the power curve 1 h before the point in time to be predicted.

MASM is a technical indicator that utilizes the aggregation and separation conditions between short-term averages and raw data combined with the time series characteristics of the averages themselves to investigate and judge the highs and lows of the prediction object (Li, 2013). The principle of MASM is to use the EMA that characterizes the short-term trend of the raw data and to compute the PMSF of the current instantaneous rate of change of the EMA. The PMSF can better project the inflection point of the trend after the comprehensive evaluation of the mutation sensitivity. The specific steps of MASM are as follows:

Find the N-day smoothed moving average X of t if X' is the N-1 day smoothed moving average:

$$X = \mathrm{EMA}(t, N) = \frac{(N-1) \times X' + t}{N} \tag{1}$$

Where: t is the current time point; X is the N-day smoothed moving average of the time series at moment t; X' is the N-1 day smoothed moving average of the time series.

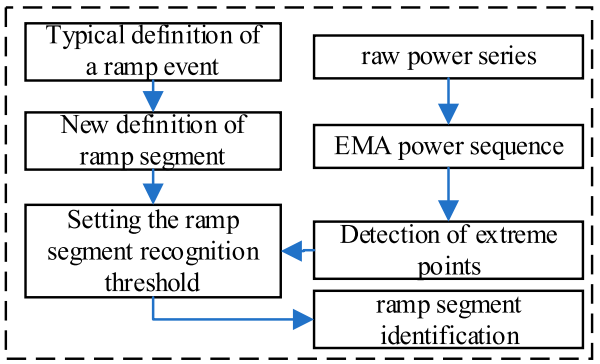


FIGURE 3
Basic flow of ramp segment recognition.

The EMA curve obtained above is smoothed by the Gaussian window method, and the rate of change of each moment in the time sequence is further calculated as PMSF. The specific calculation is as follows:

$$\mathrm{PMSF} = \frac{X_{\mathrm{smooth}}(t) - X_{\mathrm{smooth}}(t - \Delta t)}{\Delta t}, \quad \Delta t \to 0 \tag{2}$$

Where: $X_{\mathrm{smooth}}(t)$ is the smoothed EMA value at the time $t - \Delta t$; $X_{\mathrm{smooth}}(t - \Delta t)$ is the smoothed EMA value at time $t - \Delta t$. The effect of EMA and PMSF applications is shown in Figure 2. The EMA and PMSF are calculated using Equations 1, 2.

4 Ramp segment identification

Considering that a ramp event is a large change in wind power over a short period, the wind power ramp event can be redefined by the ramp amplitude and ramp rate. In this paper, we will first find the extreme points of historical wind power sequences, and analyze and identify the ramp events based on the sequence of extreme points to avoid the identification of ramps under different definition criteria.

The current wind power ramp is generally studied as an "event", and the complete ramp event consists of multiple ramp segments, so this paper will take the "ramp segment" as an object to study, and put forward a new approach to identify the ramp segment, the basic idea is shown in Figure 3.

As can be seen in Figure 3, on the one hand, the original wind power sequence is extracted from the extreme point to find out the extreme sequence; on the other hand, the new definition is determined by the typical definition of ramp events; the magnitude threshold and rate threshold are set in combination with the above two aspects to identify the ramp segment; finally, the feature analysis is carried out to determine the ramp segments in a specific region.

4.1 Extreme extraction process

The extreme extraction method achieves the effect of feature extraction by extracting the extreme values of the original sequence,

by searching the local extreme points of the numerical sequence. Assuming that the original information matrix is $\boldsymbol{X}$, $\boldsymbol{X}$ can be expressed as.

$$\begin{cases} \boldsymbol{X} = \begin{bmatrix} T_1, ..., T_i, ..., T_m \\ P_1, ..., P_i, ..., p_m \end{bmatrix}^{\mathrm{T}} \\ i = 1, ..., m \end{cases} \tag{3}$$

Where: $\boldsymbol{X}$ is the original data matrix; T_1, T_i, and T_m are the 1st, I, and mth data moments, respectively; P_1, P_i and p_m are the 1st, i, and mth power data, respectively; i is the counting point; m is the amount of original data. Extracting the extreme sequence from $\boldsymbol{X}$, the specific process is as follows:

1) Initialize the beginning and end of $\boldsymbol{E}$ as the beginning and end of $\boldsymbol{X}$:

$$\begin{cases} \boldsymbol{E}(1,1) = T_1; \ \boldsymbol{E}(1,2) = P_1 \\ \boldsymbol{E}(d,1) = T_m; \ \boldsymbol{E}(d,2) = P_m \end{cases} \tag{4}$$

Where: $\boldsymbol{E}$ is the extreme matrix; d is the number of extreme points.

2) Give the discriminant of the extreme point:

$$\begin{cases} ①: P_{i-1} > P_i < P_{i+1} \\ ②: P_{i-1} < P_i > P_{i+1} \end{cases} \tag{5}$$

If condition ① is met, it is a point of minimal value; if ② is met, it is a point of maximum value. Store T_i、 P_i to $\boldsymbol{E}$ at the extreme point.

3) Correction for the beginning of the polar matrix.

$$S_{\mathrm{bin}} = \frac{\boldsymbol{E}(2,2) - \boldsymbol{E}(3,2)}{\boldsymbol{E}(2,1) - \boldsymbol{E}(3,1)} \tag{6}$$

$$L_{\mathrm{bin}} = S_{\mathrm{bin}} \left(\boldsymbol{E}(1,1) - \boldsymbol{E}(2,1)\right) + \boldsymbol{E}(2,2) \tag{7}$$

Where: S_{bin} is the magnitude of the ramp, from the start to the end; L_{bin} is the first and last of the positive correction. The first and last of $\boldsymbol{E}$ are corrected using (Equations 6, 7).

(4) Extreme matrix terminal correction.

$$S_{\mathrm{end}} = \frac{\boldsymbol{E}(d-1,2) - \boldsymbol{E}(d-2,2)}{\boldsymbol{E}(d-1,1) - \boldsymbol{E}(d-2,1)} \tag{8}$$

$$L_{\mathrm{end}} = S_{\mathrm{end}} \left(\boldsymbol{E}(d,1) - \boldsymbol{E}(d-1,1)\right) + \boldsymbol{E}(d-1,2) \tag{9}$$

Where: S_{end} is the terminal ramp; L_{end} is the terminal correction value. The terminal of E is corrected using (Equations 8, 9).

The above content provides a detailed introduction to the basic principles of the extreme extraction method and practical operation steps. We utilize the wind power data measured at a wind farm to verify the reliability of the extreme extraction method every 15 min, and the verification results are shown in Figure 2C.

The solid line in Figure 2C is the characteristic line of the extracted extreme points. From Figure 2C, it can be seen that the extreme extraction method can effectively extract the extreme points in the power series, and less extreme data can be used in the presentation of the change characteristics of the original series, to achieve the effect of data compression and achieve the purpose of feature extraction.

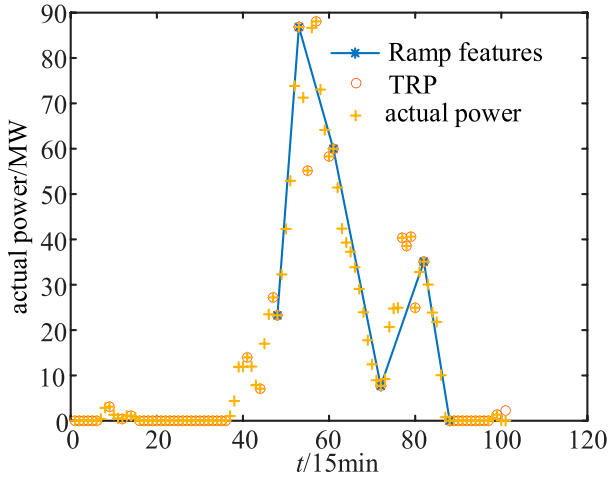


FIGURE 4
Ramp segments identified according to Definition I.

To facilitate the analysis of the following article, the extreme points are called temporary ramp points (TRP), and the $\boldsymbol{E}$ is called TRP series $\boldsymbol{Y}$, that is:

$$\begin{cases} \boldsymbol{Y} = \begin{bmatrix} T_1^{\mathrm{Y}}, ..., T_j^{\mathrm{Y}}, ..., T_n^{\mathrm{Y}} \\ P_1^{\mathrm{Y}}, ..., P_j^{\mathrm{Y}}, ..., P_n^{\mathrm{Y}} \end{bmatrix}^{\mathrm{T}} \\ j = 1, .., n \end{cases} \tag{10}$$

Where: $\boldsymbol{Y}$ is the TRP matrix; T_j^{Y}、 P_j^{Y} are the TRP moments with power; j is the number of counts of temporary ramps; n is the number of temporary ramps.

4.2 Definition of the ramp segment and threshold setting

4.2.1 Definition of a ramp segment

1) The classical definition of a ramp event.

Literature (Zhang et al., 2018; Freedman et al., 2008; Cutler et al., 2011) summarizes several typical definitions of ramp events.

Definition I. The condition is met if the difference between the power at the cutoff time and the power at the start time exceeds a set threshold λ within a given period $[t, t+\Delta t]$.

$$|P_{t+\Delta t} - P_t| > \lambda \tag{11}$$

Then the power ramp event is considered to occur in this time frame, where λ is the threshold of the ramp amplitude.

The literature (Zhang et al., 2018) recommends that a change in amplitude greater than about 15%–20% of the total installed capacity is recognized as a ramp event. The threshold of ramp magnitude for wind farms is used for the test, which results in an average value of about 30–40 MW. Figure 4 shows the ramp identification plot according to Definition I (Equation 11), taking λ = 35 MW.

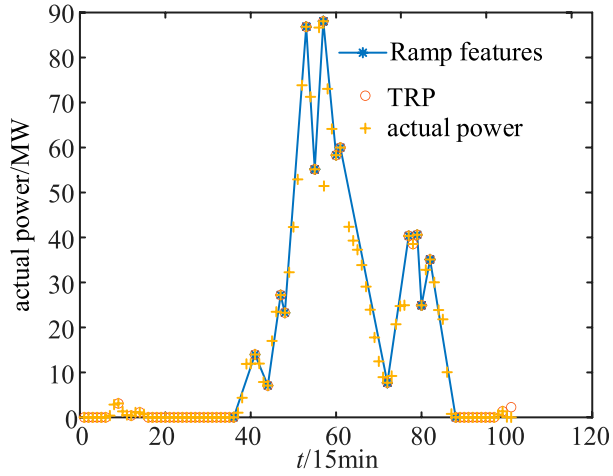


FIGURE 5
Climbing segments identified according to Definition III.

From Figure 4, this definition recognizes simple ramp events in the power sequence. However, only the ramp amplitude is considered, and the ramp rate (the change characteristics of the power in the ramp process) is not considered, resulting in the loss of power characteristics. In addition, the ramp amplitude threshold set only based on the installed wind farm cannot effectively reflect the actual power amplitude change of the wind farm, which is easy to causes the omission of identification.

Definition II. Firstly, a time range $[t, t+\Delta t]$ is circled, and then the maximum and minimum values are searched in this time range to find their extreme difference which is larger than λ, that is

$$\max\left(P_{[t,t+\Delta t]}\right) - \min\left(P_{[t,t+\Delta t]}\right) > \lambda \tag{12}$$

Then the power ramp event is considered to occur in this time interval. Definition II considers the power amplitude in the time interval, while the rate of change is not characterized.

Definition III. Circle a certain time range $[t, t+\Delta t]$, and when the power rate is greater than the value of β, it is the ramp rate in that time range:

$$\frac{|P_{t+\Delta t} - P_t|}{\Delta t} > \beta \tag{13}$$

Then the power ramp event is considered to occur within this time frame, where the β is the threshold for ramp rate. This definition is simultaneously able to determine the up-ramping and down-ramping situations. Equation 13 defines an up-ramp event when $P_t < P_{t+\Delta t}$ and a down-ramp event when $P_t > P_{t+\Delta t}$.

According to the literature (Truewind, 2008; Cutler et al., 2011), only when the power change of the wind farm reaches at least 50% of the installed capacity within 4 h is recognized as a ramp event, so the corresponding rate threshold can be calculated according to 0.417 MW/min. Combined with Equation 13, the ramp diagram identified by Definition III is drawn, as shown in Figure 5.

As can be seen in Figure 5, although the dynamics can be accurately depicted based on the ramp rate, the information redundancy of the ramp events is also increased by the extraction points with small change amplitude, leading to an unclear identification of the ramp events.

The definition of the above ramp event method identifies different results, which is not popularized in practical applications. And for the ramp events with complex processes and long periods, they are often interspersed with non-ramp intervals which cause ramp misrecognition. In the following segment, we will take a segmented approach, and based on the definition of a typical ramp event, we will define the ramp segment, and discuss and find the method of setting the ramp threshold.

2) A new type of ramp segment definition.

As can be seen from Figures 4, 5, it is difficult to recognize the complex process of ramp events by using Definitions I–III alone. In this paper, the ramp event is segmented, and Definitions I–III are combined and refined to redefine the ramp segment by combining (Equations 3–9) with the $\boldsymbol{Y}$, as follows:

$$\begin{cases} \text{ramp points: } \left|P^{\mathrm{Y}}_{j+1} - P^{\mathrm{Y}}_{j}\right| > \lambda\,(\mathrm{C1})\text{and} \\ \beta_{\max} > \dfrac{\left|P^{\mathrm{Y}}_{j+1} - P^{\mathrm{Y}}_{j}\right|}{T^{\mathrm{Y}}_{j+1} - T^{\mathrm{Y}}_{j}} > \beta\,(\mathrm{C2}) \\ \text{stationary point: } \left|P^{\mathrm{Y}}_{j+1} - P^{\mathrm{Y}}_{j}\right| < \lambda \text{ or } \dfrac{\left|P^{\mathrm{Y}}_{j+1} - P^{\mathrm{Y}}_{j}\right|}{T^{\mathrm{Y}}_{j+1} - T^{\mathrm{Y}}_{j}} < \beta \end{cases} \tag{14}$$

where $\beta_{\max}$ is the highest value of the rate of change of the incoming power determined by the wind farm installation and the grid, with the specific reference values shown in Table 1.

Equation 14 makes a limitation on the ramp amplitude (C1) and ramp rate (C2); when the C1 is available, the power rate between TRPs (Equation 10) is examined; when the C1 and C2 are satisfied at the same time, it is determined that the ramping has occurred between the TRPs, and the P^{Z}_{q} is judged to be a ramp point, see the following ramp point matrix $\boldsymbol{Z}$:

$$\begin{cases} \boldsymbol{Z} = \begin{bmatrix} T^{\mathrm{Z}}_{1}, \ldots, T^{\mathrm{Z}}_{q}, \ldots, T^{\mathrm{Z}}_{r} \\ P^{\mathrm{Z}}_{1}, \ldots, P^{\mathrm{Z}}_{q}, \ldots, P^{\mathrm{Z}}_{r} \end{bmatrix}^{\mathrm{T}} \\ q = 1, .., r-1 \end{cases} \tag{15}$$

Where: $\boldsymbol{Z}$ is the ramp point matrix; T^{Z}、P^{Z} are the moment and power of the ramp point; q is the number of counts of ramp points; r is the number of ramp points.

The combination of condition I and condition II (Equation 12) makes the information redundancy better in complex ramp segments. Since $T^{\mathrm{Y}}_{j+1}, T^{\mathrm{Y}}_{j}$ are not a fixed time range, therefore, λ and β do not constitute a fixed mathematical relationship, but two independent conditional thresholds, and the setting between λ and β is discussed on this basis.

Equation 14 introduces the concept of stationary point (SP), defines the TRP with ramp amplitude less than λ or ramp rate less than β as a stationary point, and replaces the successive stationary points with a horizontal line, whose value is equal to the value of the starting stationary point. By introducing the concept of stationary point, clear statistics of the ramp period, reducing the redundancy of information in the ramp segment, and reducing the statistical error

TABLE 1 Recommended value of maximum power rate change of wind farm.

Installed capacity of wind farms/MW	Maximum change in 10 min/MW	Maximum change in 1 min/MW
<30	20	6
30–150	Cap/1.5	Cap/5
>150	100	30

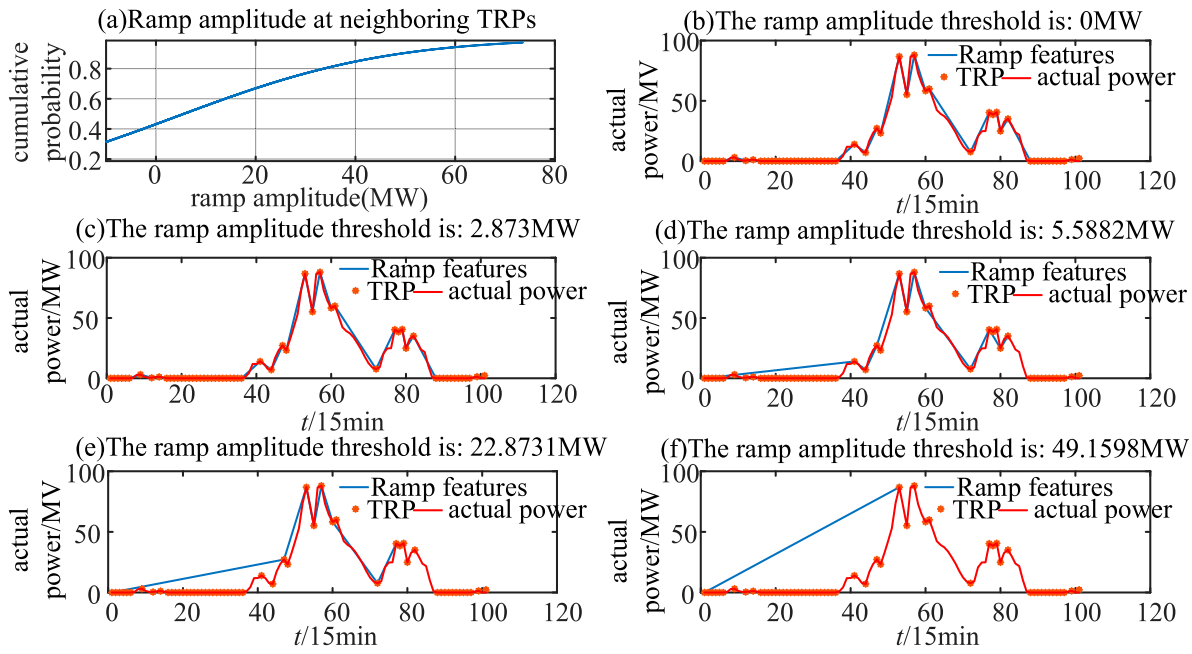


FIGURE 6
Ramp diagram under different ramp thresholds. **(A)** Ramp amplitude at neighboring TRPs. **(B)** The ramp amplitude threshold is 0 MW. **(C)** The ramp amplitude threshold is: 2.873 MW. **(D)** The ramp amplitude threshold is: 5.5882 MW. **(E)** The ramp amplitude threshold is: 22.8731 MW. **(F)** The ramp amplitude threshold is: 49.1598 MW.

of ramp duration, we can effectively distinguish and recognize the calm and slow zones during power ramp.

4.2.2 Discussion of ramp threshold setting

1) Setting of β

C2 of Equation (14) already specifies the method for calculating the ramp rate. As shown in Figure 6A, the ramp rates of adjacent TRP are calculated and counted, and the confidence intervals of the ramp rates are analyzed by their cumulative probability distribution plots. Figure 6A shows that more than 90% of the ramp rates are less than βmax, indicating that the majority of the ramp rates can meet the grid connection

Requirements, and more than 80% of the rates are greater than 0.417 MW/min. The ramp rate threshold for this wind farm is 0.4 MW/min, which can be obtained according to the recommendation of Literature (Huang et al., 2016), which indirectly verifies the validity of determining the ramp threshold by the method of mathematical statistics. According to the definition of Equation 14, when the rate of change of power in a certain time range is greater than a certain threshold value, it can be determined that the ramp may occur in this period. According to the mathematical and statistical results in Figure 6A, more than 80% of the power changes are contained in cases where the absolute value of the power rate of change is greater than or equal to 0.417 MW/min, which is consistent with today's prevailing view of hill-ramp rate threshold setting.

2) *Setting* of λ.

To address the problem of setting λ (Equations 11, 12), this paper utilizes mathematical statistics to determine the ramp magnitude threshold by counting the power magnitude changes between sequences of TRP, as shown in Figures 6B–F for β = 0.4314 MW/min, λ = 2.875 MW, 5.588 MW, 22.873 MW, and 45.1598 MW. From the figure, it can be seen that different ramp amplitudes can recognize different ramp processes, and lower amplitude thresholds can have a good description of the ramp process, but it is impossible to exclude the interference of small fluctuations in power, which results in wrong recognition. With higher amplitude thresholds, small fluctuations in power can be excluded and the ramp process can be depicted with fewer points. Setting different ramp amplitude thresholds and combining the actual dynamic changes in the wind field, is a method to select the optimal ramp amplitude threshold by choosing the probability interval of changes with different amplitudes.

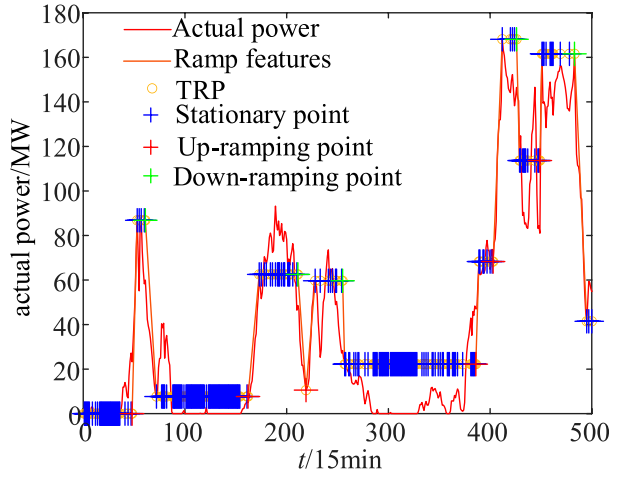


FIGURE 7
Wind power ramp diagram based on extreme point extraction.

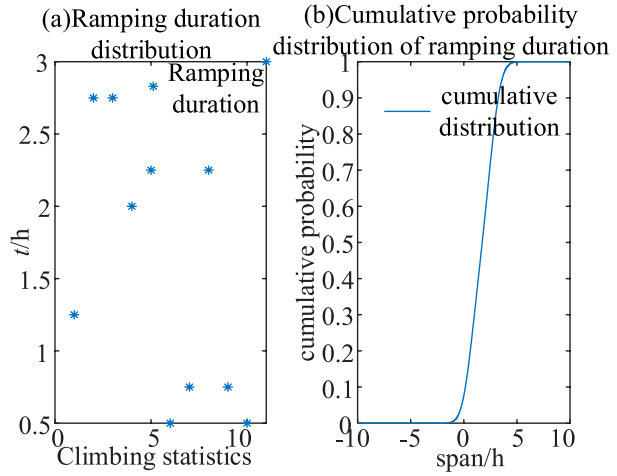


FIGURE 8
Ramp duration chart. **(A)** Ramping duration distribution. **(B)**. Cumulative probability distribution of ramping duration.

The determination of ramp amplitude thresholds using the mathematical and statistical method has an obvious advantage in that the thresholds can be set flexibly, and this threshold setting is more prominent compared to the traditional method. The magnitude of historical power changes is summarized through statistics and analysis of historical data of a specific wind field to set the ramp magnitude threshold flexibly, thus providing flexible identification of ramp events.

4.3 Ramp segment identification based on extreme point extraction

Through the redefinition of the ramp segment and the discussion of the ramp threshold, combined with the experience of the selection of the ramp threshold in the typical definition of ramp, the uphill point is identified as an ascent point when the ramp rate is greater than or equal to the value of β, the typical definition of the ramp rate is less than the value of β, and the typical definition of the ramp rate is greater than or equal to the value of β as the characteristic ramp identification schematic as shown in Figure 7. And the time of ascent was counted according to the ascent schematic, and the results are shown in Figure 8. The formula is as Equation 16 to discriminate the point of up and down the ramp, and the formula is calculated as Equation 17 the duration of ramp.

$$\begin{cases} \text{up}-\text{ramping}: \dfrac{P^{Z}_{q+1}-P^{Z}_{q}}{T^{Z}_{q+1}-T^{Z}_{q}} > \beta \\ \text{down}-\text{ramping}: \dfrac{P^{Z}_{q+1}-P^{Z}_{q}}{T^{Z}_{q+1}-T^{Z}_{q}} < -\beta \end{cases} \tag{16}$$

$$T = \begin{cases} T^{Z}_{q+1} - T^{Z}_{q+1} \\ T_{FSP} - T^{Z}_{q+1} \end{cases} \tag{17}$$

where T_{FSP} is the moment of the starting stationing point.

Figure 8A shows the distribution of ramp length, 12 ramp segments were identified, the shortest segment was 0.5 h, the longest segment was 3 h, and the average time of each segment was 1.8 h. Figure 8B shows the cumulative distribution of continuous time ramp, with more than 90% of the continuous time counted in the area within 2.234 h.

Combined with the previous analysis, the basic flow of this algorithm is given as follows: based on the extraction of extreme value points, the identification of ramp segments is completed.

Step 1. Extract the extreme point of the historical actual power sequence, according to the extreme value extraction method and Equations 3–9 to get the extreme value series ***E***, also known as the TRP series ***Y***.

Step 2. the amplitude change of power, according to C1 of Equation 14 and the TRP sequence calculation statistics, analyzes the cumulative probability distribution of amplitude change by calculation, sets the ramp amplitude threshold limit, and finds out Equation 14.

Step 3. sets the ramp rate threshold and calculates and analyzes the power rate of C2 of Equation 14 and the TRP sequence by calculating and analyzing the cumulative probability distribution of the rate of change.

Step 4. Identify the stationary point based on the stationary point identification condition in Equation 14 Identify the stationary point based on Equations 15, 16 based on the parameters determined by STEP2 and STEP3, and identify the upper and lower ramps based on Equations 15, 17 to count the length of the ramp time.

5 Segmentation prediction algorithm

To improve the performance of wind power prediction under the full-time period and make the model fit the typical characteristics of various meteorological types, this paper

proposes a time-series segment prediction algorithm that combines point prediction and probabilistic prediction. Based on the extraction results of meteorological periods in the ramp segment in Section 3, the time series is divided into a fluctuation segment and a low output leveling segment, corresponding to the probability prediction and point prediction respectively. This method helps to improve the reasonableness and readability of the prediction results, and reduces the amount of modeling operations, and improves the computational efficiency.

5.1 LightGBM-LSTM based point prediction method

In practical applications, full-time power prediction challenges a single prediction model. In this paper, a combined LightGBM-LSTM prediction method is proposed to improve the problem of insufficient prediction accuracy in specific scenarios. Massive feature data are used as inputs to the LSTM network and LightGBM prediction model, respectively; meanwhile, to improve the prediction accuracy of the LSTM neural network, the preliminary prediction result of the LightGBM model is input into the LSTM network as one of the features. This combined prediction model can combine the respective features of the above 2 models, which can not only explore the intrinsic connection between multi-feature data but also avoid the bad influence on the prediction accuracy due to over-fitting.

5.1.1 LightGBM model

LightGBM is a framework for implementing the gradient boosting decision tree algorithm. The training speed is faster, the memory consumption is lower, the accuracy is better, the support of distributed, and the fast processing of massive data can be performed (Chen et al., 2021). The main improvements of the LightGBM model include the histogram algorithm and the Leaf-Wise Reading Strategy (Ju et al., 2019). Among them, the former can be substantially reduced in terms of memory usage; the latter can grow deeper decision trees for better prediction accuracy with the same number of breaks. In addition, LightGBM supports category characteristics that do not need to be transformed (e.g., whether it is weather on a ramp road) and incorporates decision rules for category characteristics in the decision tree algorithm.

5.1.2 LSTM

LSTM is a special kind of Recurrent Neural Network architecture (Nguyen. et al., 2024), which can solve the problem of modeling time-series data of integrated drives. LSTM can effectively deal with the long-time dependency relationship and introduces the "memory unit" and gating mechanism. The structure of LSTM is shown in Figure 9.

Figure 9 shows the input X_t, the output H_t, the memory cell state C_t, and the candidate memory cell state $\tilde{C}_t$ of the LSTM at the t time.

The C_t is the main memory unit responsible for storing and transmitting information in the LSTM. It is similar to the hidden layer in a traditional neural network, but it gets updated with information at every step. This design allows it to maintain long-term memory.

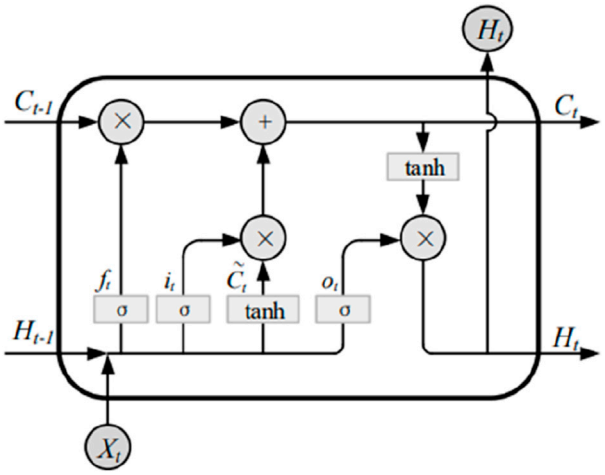


FIGURE 9
Model structure of LSTM.

The input information for each step of the LSTM contains the C_{t-1}, H_{t-1} of the previous step and the X_t of the current step, which allows it to preserve long-term dependencies over the entire time series. The core of the LSTM is a 3-gate structure, the oblivion gate, the input gate, and the output gate, whose outputs are f_t, i_t and o_t, respectively.

The oblivion gate decides which information of C_{t-1} from the previous step is to be forgotten. Its uses a Sigmoid function to obtain the f_t. It is a vector in the range [0, 1] which is used to control how much of C_{t-1} is forgotten.

The computational expression for f_t is:

$$f_t = \sigma(w_f[H_{t-1}, X_t] + b_f) \tag{18}$$

Where: σ is the activation function of the gate structure; w_f, b_f are the weight and bias of the oblivion gate.

The input gate controls the input information of the current step. It consists of two parts: one part uses a tanh function to filter valid information from X_t as $\tilde{C}_t$; the other part uses a Sigmoid function to obtain i_t, which is used to control the degree of validity of the candidate memory cells.

The computational expressions for the i_t, $\tilde{C}_t$ and C_t are respectively:

$$i_t = \sigma(w_i[H_{t-1}, X_t] + b_i) \tag{19}$$

$$\tilde{c}_t = \tanh(w_c[H_{t-1}, X_t] + b_c) \tag{20}$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{c}_t \tag{21}$$

where: w_i, b_i are the input gate weights and bias; w_c, b_c are the weight and bias of the candidate memory cells; $\otimes$ is the element-by-element product.

The output gate determines the output information of the current time step, which uses the Sigmoid function to obtain o_t; o_t and C_t together determine the H_t of the neuron at the current time step.

The computational expressions for the o_t and the H_t are:

$$o_t = \sigma(w_o[H_{t-1}, X_t] + b_o) \tag{22}$$

$$H_t = o_t \otimes \tanh(C_t) \tag{23}$$

where: w_o, b_o are the output gate weight and bias. In summary, the basic model of LSTM is formed by Equations 18–23.

5.2 Probabilistic prediction methods based on temporal pattern classification

5.2.1 Convolutional neural networks

CNN is a feed-forward neural network type of deep learning model, due to its properties of extracting spatial features commonly used in various data analysis, computer vision, natural language processing, and other fields, so a convolutional layer is set up in the data to extract spatially correlated features. Especially for the characteristics of multi-dimensional extraction of data, to reduce the complexity of the problem; the pooling layer is designed to reduce the dimensionality and number of data so that it can reduce a lot of features that need to be operated to improve the learning efficiency; with the development of deep learning, more and more forms of convolution can be exercised to improve the effect of residual convolution, Alexnet and other convolution of different functions.

5.2.2 Bidirectional gated recurrent unit (Bi-GRU)

GRU uses recursion to obtain global information from the input sequence, utilizes update gate and reset gate to reduce gradient dispersion, and achieves the ability to remember the sequence over time and less computational loss. The update gate determines how much previous information is currently retained at the forecast point.

$$z_t = \sigma(W_z \cdot [h_{t-1}, Fea_t] + b_z) \tag{24}$$

where: z_t is the output of the update gate; Fea_t denotes the input matrix for time step t. h_{t-1} is the hidden state of the previous time step t-1. W_z b_z are weight and base of the update gate.

Reset gate controls how much historical information should be ignored and determines whether the storage unit removes unnecessary detection features. Described as

$$r_t = \sigma(W_r \cdot [h_{t-1}, Fea_t] + b_r) \tag{25}$$

where: r_t is the output of the reset gate; W_r, b_r are weight and base of the reset gate.

Effective forecasting models need to extract implicit features and complex changes in serial data. However, GRU can only extract information from the forward direction, while ignoring the valuable information in the backward time series data. Therefore the algorithmic idea of Bi-GRU is proposed, in which the Bi-GRU layer in the encoder consists of two independent GRU networks as shown in Figure 2. They are interconnected at adjacent depths to ensure that the hidden layer state at the previous depth can be transferred to the next hidden state in one direction and features can be extracted from both directions. Bi-GRU can be represented as

$$h_T = F\left(\vec{L}_t, \overleftarrow{L}_t\right) \tag{26}$$

where: $\vec{L}_t$, $\overleftarrow{L}_t$ are the hidden states of the forward and backward GRUs. F denotes how the outputs of the two directions are combined, e.g., multiplication function, averaging function, summation function, etc. In summary, the basic unit model of Bi-GRU is constituted by Equations 24–26.

6 Calculation validation

6.1 Description of experimental data

Selected for this test sample is information from a wind farm in the country. The data and information of this experiment include the annual output power of this wind farm in 2021 as well as a variety of meteorological factors during the same period. All the above data intervals are 15 min. To verify the effectiveness and superior performance of the algorithm of this paper for the weather conditions of the ramp segment, 10% of the wind farm data of variable meteorological scenarios containing the weather of the ramp segment are specially selected as the validation dataset.

6.2 Evaluation indicators

The article evaluates the prediction performance in terms of both deterministic and uncertainty prediction metrics. Among the deterministic prediction evaluation metrics include relative root mean square error (RRMSE) and mean absolute percentage error (MAPE).

$$\mathrm{RRMSE} = \left[n\sqrt{\frac{1}{n}\sum_{i=1}^{n}(f - f_{\mathrm{his}})^2}\right] \Big/ \left(\sum_{i=1}^{n} f_{\mathrm{his}}\right) \tag{27}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{f - f_{\mathrm{his}}}{f_{\mathrm{his}}}\right| \tag{28}$$

Where: n is the total number of prediction samples; f is the prediction value; f_{his} is the true power value.

Uncertainty prediction evaluation metrics include prediction interval coverage percentage (PICP) and prediction interval average width (PINAW) (Ushakov and Ushakov, 2012).

$$R_{\mathrm{PICP}} = \frac{1}{W}\sum_{w=1}^{W} k_{\mathrm{wa}} \tag{29}$$

$$R_{\mathrm{PINAW}} = \frac{1}{T}\sum_{t=1}^{T} U(x_t) - L(x_t) \tag{30}$$

Where: R_{PICP} is the PICP value; W is the point to be predicted and is taken as 250 in this paper; k_{wa} is a Boolean quantity, k_{wa} = 1 means that the actual power value of the point to be measured falls within the prediction interval at the given confidence level. R_{PINAW} is the PINAW value; T is the time-range prediction; $U(x_t)$, $L(x_t)$ are the upper and lower power predictions. A smaller PINAW corresponds to a better prediction when the PICPs are equal.

6.3 Validation of point prediction results

The LIGHTGBM model in the MATLAB platform is used to call the LIGHTGBM machine learning library, the number of weak back trees is 200, the number of leaves is 50, the learning rate is 0.05, and

TABLE 2 Presents the recommended maximum power rate change value for wind farms.

Model	RRMSE	MAPE
LSTM	12.475	10.843
LightGBM	8.857	6.934
LightGBM-LSTM	6.734	4.174

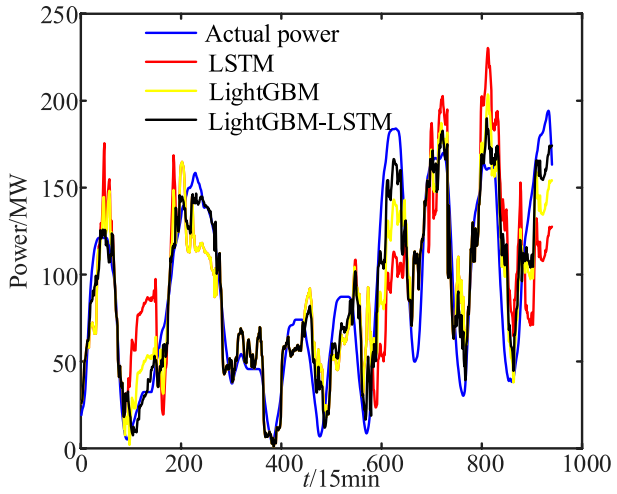


FIGURE 10
Comparison of point prediction performance.

the number of iterations is 3,000; the short-term prediction model of wind power based on the LSTM model constructed by using the KERAS framework. The initialization parameters of the prediction model are: the number of nodes in the hidden layer of the model network and the learning rate of the weights are determined by the IGWO algorithm, the number of model iterations is 15, and the activation function of the LSTM model uses the Sigmoid function.

The data for the training set of the model is the first 80% of the full year 2021 data, while the data for the test set is the last 20%. To observe the prediction effect over a longer time interval, the 2021 data is subjected to several sets of short-term 60-day rolling forecasts with a forecast window of 10 days, i.e., the power changes are predicted 10 days in advance for the next 10 days. As described in 5.1, the algorithm especially selects the power-abrupt time-series segments containing the weather of the ramp segment as the validation dataset and predicts the wind power under the variable meteorological scenarios using LightGBM (Zheng and Kusiak, 2009; Zareipour et al., 2011b), LSTM (Greaves et al., 2009), and LightGBM-LSTM, respectively, and the prediction results of the algorithms are shown in Table 2. The resultant curves are predicted in different ways, as shown in Figure 10.

Comparing the prediction curves in Figure 10, it can be seen that the prediction effect of LightGBM-LSTM algorithm is closer to the real value. And it can be further seen from the prediction errors (RRMSE and MAPE were calculated using Equations 27–28) in Table 2 that it has better performance in both RRMSE and MAPE, indicating that the model not only has high prediction accuracy, but also can capture the features of the data well and give predictions

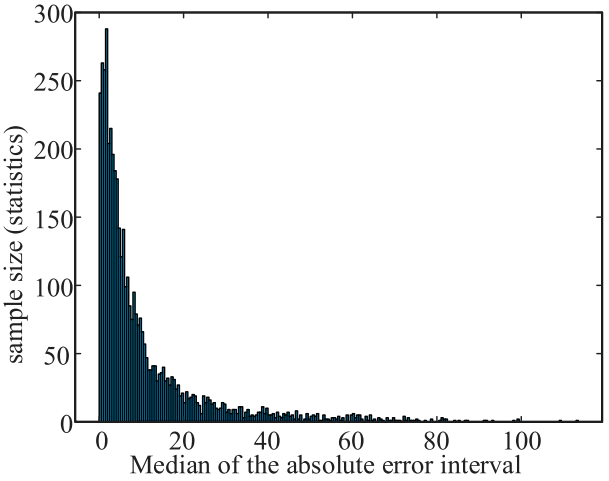


FIGURE 11
Error distribution of probabilistic predictions.

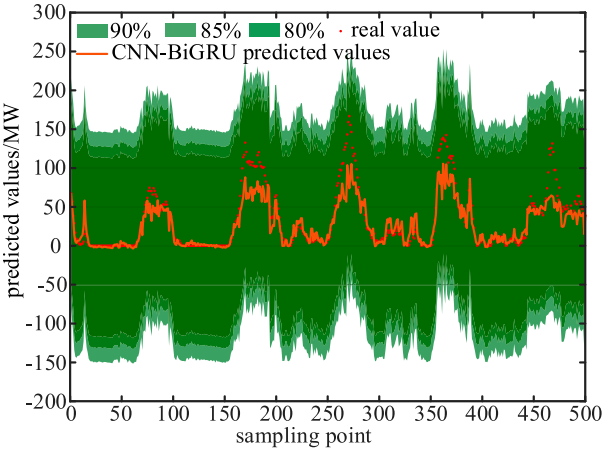


FIGURE 12
Probability interval prediction effect diagram.

close to the true value. Meanwhile it has better prediction stability and generalization ability to prevent overfitting. The improved LightGBM-LSTM algorithm combines the respective features of LSTM and LightGBM, which can not only mine the intrinsic connection between multi-featured data, but also avoid the adverse effects of overfitting, and improve the prediction accuracy of the method by combining the prediction strategies. The LightGBM model has an obvious advantage in prediction speed, but its training process is susceptible to overfitting, which results in a lower prediction accuracy is low. The LSTM network in the combined prediction model adds LightGBM as one of the input feature tensors, and thus has a better prediction effect.

6.4 Analysis of inter-area prediction results

The data of a wind farm with an installed capacity of 201 MW in China is selected for example analysis. Figure 11 shows the

distribution of prediction error, and Figure 12 shows the prediction effect of a certain 5-day period in 2021, which shows that the interval prediction model can closely follow the trend of the wind power series under the same confidence level, and obtains a narrower average bandwidth and a higher interval coverage. It can provide more accurate forecast information for decision-makers.

The probabilistic prediction effect is shown in Figure 12, where the probabilistic prediction performance is portrayed through the PICP (Equation 29) and PINAW (Equation 30) evaluation metrics, respectively. It can be seen that the actual power timeseries curve is surrounded by the estimated confidence interval. According to the performance effect of prediction, the model is better able to extract and recognize the transitive time period timing pattern. Obviously, for the probabilistic prediction time period, the prediction intervals of lower confidence intervals are surrounded by higher confidence intervals, which effectively avoids the intersection of quartiles and proves that the power prediction method proposed in this paper has a good comprehensive performance. Further, the CNN-BiGRU-KDE probabilistic prediction method proposed in this paper shows certain accuracy advantages in the 80%, 85%, and 90% confidence intervals, which further demonstrates the effectiveness and superiority of the method proposed in this paper. Meanwhile, it can obtain interval coverage greater than the preset confidence level at different confidence levels. And the probabilistic prediction method based on CNN-BiGRU-KDE not only produces favorable bias at all confidence levels, but also has high reliability.

Based on the above analysis of the deterministic prediction results, it is easy to see that the LightGBM-GRU algorithm adopted in this paper shows a high prediction effect in the small fluctuation and gentle power period. However, for the weather segments with violent fluctuations and ramp segments, there is still a certain error. Especially, it is more obvious in the period of frequent large waves, so the point prediction-probability interval prediction segmentation method proposed in this paper can better make up for the above shortcomings. On the other hand, the probabilistic prediction results are shown as the upper and lower bounds of the confidence interval and the probability density distribution, and the accuracy of the point prediction is sufficient to support the actual demand in the non-ramp weather segments. This is a further manifestation of the rationality of the segmented prediction strategy proposed in this paper. As can be seen from the presentation of the segmented prediction results in Figure 12 above, the actual power time series curve is surrounded by the estimated confidence interval. According to the predicted performance results, the model can extract and recognize the ramp segment period timing patterns better. The prediction intervals of lower confidence intervals are surrounded by higher confidence intervals, which effectively circumvents the interquartile crossover against the probabilistic prediction periods, proving that the comprehensive performance of the power prediction approach mentioned in this paper is good.

7 Conclusion

This paper proposes a short-term wind power segmentation prediction method based on the identification of ramp segment periods for the phenomenon of sudden power change in a short period of time under climbing segment meteorology. Compared with the existing methods, the proposed method improves the efficiency and accuracy of fluctuating period extraction in the ramp segment through the adaptive turning time period identification method based on local feature distribution; the improved LightGBM-LSTM algorithm can not only mine the intrinsic connection between multi-feature data, but also avoid the adverse effect of overfitting, and improve the prediction accuracy through the combination of prediction strategies; the proposed CNN-BiGRU-KDE probabilistic prediction method shows good prediction performance at the specified confidence level; by proposing a segmented prediction method based on the temporal pattern, it overcomes the influence of the variability of the power temporal features on the prediction results under the meteorological model, and significantly improves the model prediction performance. In summary, the prediction method proposed in this paper has good prediction accuracy in the full time period including the ramp segment weather, and has good generalization performance, which provides a certain useful supplement for the research in the field of ultra-short-term wind power prediction.

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

CY: Writing–review and editing, Writing–original draft, Methodology, Investigation, Funding acquisition, Formal Analysis. GW: Writing–original draft, Writing–review and editing, Methodology, Investigation, Funding acquisition, Data curation. YZ: Visualization, Resources, Writing–review and editing, Supervision. GB: Visualization, Writing–review and editing, Supervision, Funding acquisition. JW: Writing–review and editing, Supervision.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the Science and Technology Project of State Grid Gansu Electric Power Company Electric Power Science Research Institute under Grant (No. SGGSKY00WYJS2310221).

Conflict of interest

Authors CY and GW were employed by Power Dispatch Center of State Grid Gansu Electric Power Company. Author YZ was employed by Electric Power Science Research Institute of State Grid Gansu Electric Power 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.

The authors declare that this study received funding from Electric Power Scientific Research Institute of State Grid Gansu Power Company. The funder had the following involvement in the study: the collection and analysis of the data, as well as the funder's decision to submit the article for publication.

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

Cassola, F., and Burlando, M. (2012). Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output. *Appl. Energy* 99 (6), 154–166. doi:10.1016/j.apenergy.2012.03.054

Chen, W., Hu, Z., Yue, J., et al. (2021). Short-term load prediction based on combined model of long short-term memory network and light gradient boosting machine. *Automation Electr. Power Syst.* 45 (4), 91–97. (in Chinese).

Cheng, L., and Yu, T. (2019). A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems. *Int. J. Energy Res.* 43 (6), 1928–1973. doi:10.1002/er.4333

Cui, M., Sun, Y., and Ke, D. (2014). Prediction of wind power ramp based on sparse decomposition of atoms and BP neural network. *Automation Electr. Power Syst.* 38 (12), 6–11. (in Chinese).

Cui, M., Zhang, J., Wang, Q., Krishnan, V., and Hodge, B. M. (2019). A data-driven methodology for probabilistic wind power ramp forecasting. *IEEE Trans. Smart Grid* 10 (2), 1326–1338. doi:10.1109/tsg.2017.2763827

Cutler, N., Outhred, H., and Macgill, I. (2011). Final report on UNSW Project for AEMO to develop a prototype wind power forecasting tool for potential large rapid changes in wind power. *Univ. N. S. W Centre Energy Environ. Mark.*

Erdem, E., and Shi, J. (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. *Appl. Energy* 88 (4), 1405–1414. doi:10.1016/j.apenergy.2010.10.031

Ernst, B., Oakleaf, B., Ahlstrom, M., Lange, M., Moehrlen, C., Lange, B., et al. (2007). Predicting the wind. *IEEE Power Energy Mag.* 5 (6), x5–x78. doi:10.1109/mpae.2007.4317489

Ferreira, C., Gama, J., Matias, L., et al. (2011). A survey on wind power ramp forecasting. *Rep. ANL/DIS-10-13, Argonne Natl. Lab.*

Freedman, J., Markus, M., and Penc, R. (2008). *Analysis of West Texas wind plant ramp-up and ramp-down events*. New York: AWS Truewind.

Gao, Y., Qu, C., and Zhang, K. (2016). A hybrid method based on singular spectrum analysis, firefly algorithm, and BP neural network for short-term wind speed forecasting. *Energies* 9 (10), 757. doi:10.3390/en9100757

Greaves, B., Collins, J., Parkes, J., and Tindal, A. (2009). Temporal forecast uncertainty for ramp events. *Wind Eng.* 33 (33), 309–319. doi:10.1260/030952409789685681

Haoyi, X., Xiaoxia, H., and Chunli, L. (2023). Probability density forecasting of wind power based on transformer network with expectile regression and kernel density estimation. *Electronics* 12 (5), 1187. doi:10.3390/electronics12051187

Hu, Q., Zhang, S., Xie, Z., Mi, J., and Wan, J. (2014). Noise model based vv-support vector regression with its application to short-term wind speed forecasting. *Neural Netw.* 57 (9), 1–11. doi:10.1016/j.neunet.2014.05.003

Huang, Q., Wang, Z., Du, B., et al. (2016). Research on wind power grade ability prediction based on pre-decomposition combined forecasting. *Renew. Energy* 34 (12), 1847–1852. (in Chinese).

Jianhou, W., Yue, Y., Bo, Z., and Lu, H. (2024). Hybrid ultra-short-term PV power forecasting system for deterministic forecasting and uncertainty analysis. *Energy* 288, 129898. doi:10.1016/j.energy.2023.129898

Ju, Y., Sun, G., Chen, Q., Zhang, M., Zhu, H., and Rehman, M. U. (2019). A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *IEEE Access* 7, 28309–28318. doi:10.1109/access.2019.2901920

Kamath, C. (2010). "Understanding wind ramp events through analysis of historical data," in *PES transmission and distribution conference and exposition* (New Orleans, USA: IEEE), 1–6.

Kamath, C. (2011). "Associating weather conditions with ramp events in wind power generation," in Power Systems Conference and Exposition, Phoenix, AZ, USA, 20-23 March 2011 (IEEE), 1–8.

Lefeng, C., Guiyun, L., Hanqi, Hu., Wang, X., Chen, Y., Zhang, J., et al. (2020). Equilibrium analysis of general N-population multi-strategy games for generation-side long-term bidding: an evolutionary game perspective. *J. Clean. Prod.* 276, 124123. doi:10.1016/j.jclepro.2020.124123

Lefeng, C., Linfei, Y., Jianhui, W., et al. (2021). Behavioral decision-making in power demand-side response management: a multi-population evolutionary game dynamics perspective. *Int. J. Electr. Power Energy Syst.*, 129.

Lefeng, C., Yang, C., and Guiyun, L. (2022). 2PnS-EG: a general two-population n-strategy evolutionary game for strategic long-term bidding in a deregulated market under different market clearing mechanisms. *Int. J. Electr. Power and Energy Syst.* 142 (A), 108182. doi:10.1016/j.ijepes.2022.108182

Li, C. (2013). Moving average trading strategies effectiveness research. *Shanghai Jiao Tong Univ.* (in Chinese).

Liu, R., Shi, J., Sun, G., Lin, S., and Li, F. (2024). A Short-term net load hybrid forecasting method based on VW-KA and QR-CNN-GRU. *Electr. Power Syst. Res.* 232, 110384. doi:10.1016/j.epsr.2024.110384

Men, Z., Yee, E., Lien, F., Wen, D., and Chen, Y. (2016). Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. *Renew. Energy* 87 (3), 203–211. doi:10.1016/j.renene.2015.10.014

Meng, A., Chen, S., Wang, C., et al. (2021). Ultra-short-term wind power prediction based on chaotic CSO optimized temporal attention GRU model. *Power Syst. Technol.* 45 (12), 4692–4700. (in Chinese).

Nguyen., B. N., E, O., E, P., Alberti, D., Leva, S., and Duong, M. Q. (2024). "Forecasting generating power of sun tracking PV plant using long-short term memory neural network model: a case study in ninh thuan – Vietnam," in 2024 Tenth International Conference on Communications and Electronics (ICCE), Danang, Vietnam, 31 July 2024, 333–338. doi:10.1109/icce62051.2024.10634629

Ouyang, T., Huang, H., and He, Y. (2019). Ramp events forecasting based on long-term wind power prediction and correction. *IET Renew. Power Gener.* 13 (15), 2793–2801. doi:10.1049/iet-rpg.2019.0093

Ouyang, T., Zha, X., Qin, L., et al. (2017). Wind power ramp events forecast method based on similarity correction. *Proc. CSEE* 37 (2), 572–580. (in Chinese).

Potter, C., Grimit, E., and Niissen, B. (2009). Potential benefits of a dedicated probabilistic rapid ramp event forecast tool. *Power Syst. Conf. and Expo.* 107 (1), 1–5. doi:10.1109/psce.2009.4840109

State Grid (2021). State Grid Corporation of China issues "carbon peak and carbon neutrality" action plan. *State Grid Rep.*, 001. (in Chinese).

Truewind, A. W. S. (2008). AWS True wind's final report for the Alberta forecasting pilot project. *Wind Power Forecast. PILOT Proj.*

Ushakov, N. G., and Ushakov, V. G. (2012). On bandwidth selection in kernel density estimation. *J. Nonparametric Statistics* 24 (2), 419–428. doi:10.1080/10485252.2012.655734

Wang, M., Ying, F., and Nan, Q. (2024). Refined offshore wind speed prediction: leveraging a two-layer decomposition technique, gated recurrent unit, and kernel density estimation for precise point and interval forecasts. *Eng. Appl. Artif. Intell.* 133 (PE), 108435. doi:10.1016/j.engappai.2024.108435

Wang, Y., Gao, S., Jia, W., Ding, T., Zhou, Z., and Wang, Z. (2022). Data-driven distributionally robust economic dispatch for park integrated energy systems with coordination of carbon capture and storage devices and combined heat and power plants. *IET Renew. power Gener.* 16 (12), 2617–2629. doi:10.1049/rpg2.12436

Zareipour, H., Huang, D., and Rosehart, W. (2011a). *Wind power ramp events classification and forecasting: a data mining approach*. Detroit, USA: IEEE Power and Energy Society General Meeting.

Zareipour, H., Huang, D., and Rosehart, W. (2011b). *Wind power ramp events classification and forecasting: a data mining approach*. San Diego, IEEE: IEEE Power and Energy Society General Meeting, 1–3.

Zhang, D., Dai, Y., Zhang, X., et al. (2018). Review and prospect of research on wind power ramp events. *Power Syst. Technol.* 42 (6), 1783–1792. (in Chinese).

Zheng, H., and Kusiak, A. (2009). Prediction of wind farm power ramp rates: a data-mining approach. *J. Sol. Energy Eng.* 131 (3), 376–385. doi:10.1115/1.3142727

Zhou, Y., Yu, G., Liu, J., et al. (2021). Offshore wind power prediction based on improved long-term recurrent convolutional neural network. *Automation Electr. Power Syst.* 45 (3), 183–191. (in Chinese).

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 04 November 2024
DOI 10.3389/fenrg.2024.1485369

Check for updates

OPEN ACCESS

EDITED BY
Yang Yu,
Nanjing University of Posts and Telecommunications, China

REVIEWED BY
Hao Chen,
Shanghai Maritime University, China
Lefeng Cheng,
Guangzhou University, China

*CORRESPONDENCE
Chenbin Zhou,
enetwork_sz@163.com

RECEIVED 23 August 2024
ACCEPTED 18 October 2024
PUBLISHED 04 November 2024

CITATION
Jiang X, Zhou C, Pan Q, Wang L, Wu B, Xu Y, Chen K and Fu L (2024) A multi-task learning based line parameter identification method for medium-voltage distribution network. *Front. Energy Res.* 12:1485369. doi: 10.3389/fenrg.2024.1485369

COPYRIGHT
© 2024 Jiang, Zhou, Pan, Wang, Wu, Xu, Chen and Fu. 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 multi-task learning based line parameter identification method for medium-voltage distribution network

Xuebao Jiang, Chenbin Zhou*, Qi Pan, Liang Wang, Bowen Wu, Yang Xu, Kang Chen and Liudi Fu

Suzhou Power Supply Company, State Grid Jiangsu Electric Power Co., Ltd., Suzhou, China

Accurate line parameters are critical for and dispatch in distribution systems. External operating condition variations affect line parameters, reducing the accuracy of state estimation and power flow calculations. While many methods have been proposed and obtained results rather acceptable, there is room for improvement as they don't fully consider line connections in known topologies. Furthermore, inaccuracies in measurement devices and data acquisition systems can introduce noise and outliers, impacting the reliability of parameter identification. To address these challenges, we propose a line parameter identification method based on Graph Attention Networks and Multi-gate Mixture-of-Experts. The topological structure of the power grid and the capabilities of modern data acquisition equipment are utilized to capture. We also introduce a multi-task learning framework to enable joint training of parameter identification across different branches, thereby enhancing computational efficiency and accuracy. Experiments show that the GAT-MMoE model outperforms traditional methods, with notable improvements in both accuracy and robustness.

KEYWORDS

line-parameter identification, multi-task learning, mixture of experts, medium-voltage distribution system, graph attention network

1 Introduction

The rapid development of new power systems has increased the complexity of power grid operations. The integration of distributed power sources and energy storage introduces randomness and volatility, presenting new challenges for the control and operation of distribution networks. Nowadays, power grids are mutating into Smart EEPS with highly integrated cyber systems, physical systems, and social systems. Among ML, RL has strong adaptability; thus, it is applied in many aspects of Smart EEPS, such as stability control, AGC (Automatic Generation Control), VQC (Voltage Quadergy Control), OPFC (Optimal Power Flow Control) and other scenarios (Cheng and Yu., 2019). Accurate line parameters are crucial for state estimation (SE), event detection, fault analysis, and various calculations within the distribution network (Zhang et al., 2020; Shi et al., 2020).

Unlike the transmission network, where line parameters can be derived from physical or empirical formulas based on line length, resistivity, and geometric positioning,

the distribution network requires different approaches due to its radial topology and numerous feeder nodes (Wang et al., 2016; Asprou and Kyriakides, 2018; Li et al., 2018). Blueprints and planning documents can provide design parameters, but real parameters often differ due to system upgrades. As a result, traditional transmission line parameter identification (TLPI) methods struggle when applied to distribution networks. The key challenge is linking collected data to the line model. Existing parameter estimation methods can be grouped into two categories: model-driven and data-driven.

In medium voltage distribution networks, the complexity of operations and time-varying loads make it hard to build accurate mathematical models. To obtain more accurate line parameters, real-time line parameter identification can be carried out based on measurement data obtained by on-site measuring devices (Singh et al., 2018; Yu et al., 2018; Yu et al., 2019). Currently, the data used in parameter estimation mainly comes from two types of sensors: Supervisory Control and Data Acquisition (SCADA) systems and Phasor Measurement Units (PMUs). SCADA devices have been widely installed in medium voltage distribution networks, capable of collecting the amplitude of the electrical quantities but unable to obtain the phase data. PMUs can provide synchronized electrical quantities, but their high cost has limited large-scale deployment in distribution networks, failing to meet observability requirements under most conditions. Therefore, domestic and foreign scholars have conducted research on distribution network line parameter identification methods using phase-free data (Xiao et al., 2021). When PMUs are not available (Shi et al., 2024), applied a linear regression to estimate line parameters, topology, and phase labels, with nodal angles recovered via non-linear regression.

Since measurement devices characteristics can lead to outliers, it is essential to consider the DLPI problem with outliers and propose a new robust method to improve the accuracy of line parameter identification, especially under conditions involving PMU outliers and discrepancies in the accuracy of the coefficient matrix and observation data matrix. Research methods mainly focus on the least, squares method, residual sensitivity analysis, and regression methods (Zhu and Abur, 2010; Lin and Abur, 2018). A new iterative weighted least squares (WLS) method for dealing with line parameter deviations from systematic errors is also proposed, using estimates to calculate the gain matrix and prior knowledge to calculate the covariance matrix Pegoraro and his team focused on the estimation of measurement uncertainty and correction factors of D-PMUs, conducting a series of studies (Pegoraro et al., 2017; Puddu et al., 2018; Pegoraro et al., 2019a; Pegoraro et al., 2019b; Pegoraro et al., 2022). However, these methods assume widespread deployment of micro-PMUs, which limits their application. Thus, the performance of the linear regression method is limited by the incomplete configuration of measuring equipment in distribution networks. Meanwhile, as noted in Yu et al. (2019), imperfect synchronism and time interval deviations in smart meters may not ensure instant measurements for distributed generations (DGs), flexible loads, and electric vehicles with relatively dynamic behaviors.

Thanks to the development of machine learning and deep learning technologies, data-driven methods are gradually being widely applied to analyze and extract deep insights from data based on partial real-time data. In Li et al. (2022), a differential evolution algorithm is employed to identify line parameters, even when many original parameters are missing. Chen and his team find out that the integration of heuristic swarm intelligence search algorithms and AI technologies offers a significant approach to addressing the behavioral decision-making challenges (Cheng, 2020; Cheng et al., 2021; Cheng et al., 2022). Another study (Wang and Yu, 2022) develops a physics-informed graphical learning algorithm, using stochastic gradient descent to update the three-phase series resistance and reactance (Yang et al., 2022) proposed an RBFNN-MRO method combining a radial basis function neural network with multi-run optimization, which does not require synchronized phasor measure data as it uses a constant feeder parameter model over a specified short period. Other study Li et al. (2024); Yang et al. (2023) introduced a deep-shallow neural network to approximate power flow equations, employing reinforcement learning to optimize while ensuring maximal physical consistency. To reduce the influence of noise and deviation, different robust methods are used to improve accuracy. Sun et al. (2019) use convolutional neural networks (CNNs) to classify line impedance values and the results deviate from the original within 10%. Also, recent research in parameter identification focuses on overcoming limitations related to data structure, noise, and accuracy. Graph-based models (MNGAN, MFAGCN) are proposed using attention mechanism to enhance identification with non-Euclidean structures (Xia et al., 2022; Zou et al., 2024; Wang et al., 2022).

Above all, future works in DLPI should integrate physics information with deep learning methodologies. This paper introduces a multi-branch method for identifying line parameters using data from both ends of distribution lines. Addressing current limitations in identifying parameters of branched medium voltage distribution networks with topological constraints, the paper proposes a GAT-MMoE based DLPI method. This approach employs a multi-task neural network incorporating graph convolutional networks to tackle the line parameter identification problem. The graph attention network (GAT) uses an attention mechanism to learn the importance of neighboring nodes in a graph. Unlike traditional methods, where the contribution of neighbors is fixed, GAT dynamically adjusts the influence of each neighboring node based on its relevance to the target node. This leads to more accurate and nuanced feature representation. The MMoE model extracts topology features of the distribution network, with node features derived from graph attention networks and a multi-task learning model employing homoscedastic uncertainty loss.

The rest of this paper is structured as follows: Section 2 describes the modeling of the task of multi-branch line parameter identification in the distribution network, including problem formulation and the construction of the graph attention and multi-task modules. Section 3 covers the overall framework and workflow of the suggested method. Section 4 presents results and discussion, along with dataset description and comparison to alternative machine learning methods. Finally, this paper is summarized in Section 5.

2 Problem formulation

2.1 Parameter identification task and system construct

Current line parameter identification technologies face several challenges: 1) The low investment and high construction costs of smart devices impede the deployment of PMUs at each bus node, making real-time monitoring of voltage phase angle information difficult. 2) The integration of distributed photovoltaic systems on the user side causes reverse power flow and significant voltage fluctuations, making it difficult to maintain accuracy and robustness in the task of distribution line parameter identification.

The objective of DLPI is to find the mapping between node characteristics and line parameters and to identify the line resistance and reactance of each branch. Given the features of power grid branch, it is achieved using the active power, reactive power, and voltage amplitude provided by measuring equipment in medium-voltage distribution networks. The powerful learning ability of neural networks can be utilized to build a power flow model, mine constraints, and learn historical data to train the model. The power flow calculation using polar form of the nodal power Equation 1 is:

$$\begin{cases} P_i = P_{Gi} - P_{Li} = \sum_{j=1}^{n} V_i V_j \left(G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij} \right) \\ Q_i = Q_{Gi} - Q_{Li} = \sum_{j=1}^{n} V_i V_j \left(G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij} \right) \\ Z_{ij} = R_{ij} + jX_{ij} = \dfrac{G_{ij}}{G_{ij}^2 + B_{ij}^2} - j\dfrac{B_{ij}}{G_{ij}^2 + B_{ij}^2} \end{cases} \tag{1}$$

where P_i and Q_i are the active and reactive power injected into node i, respectively. P_{Gi} and Q_{Gi} are the active and reactive power from the power source at node i; P_{Li} and Q_{Li} are the active and reactive power consumed by the load at node i; G_{ij} and B_{ij} are the conductance parameters consisting of g and b between nodes i and j, respectively. V_i and V_j are the voltage amplitudes of nodes i and j, respectively; θ_{ij} is the difference in the phase angle of the voltage between nodes i and j. In the modeling stage, we use power flow equations expressed in polar coordinates to accurately represent the system's behavior. However, during the experimental phase, the results are provided in terms of impedance parameters (resistance R and reactance X) as the original IEEE test case data is given in these terms. To facilitate direct comparison with the IEEE standard data, the node admittance matrix is converted into corresponding impedance values.

Meanwhile, parameter identification can be considered as a multi-task regression problem. Considering that the phase angle difference between the two ends of each branch of the distribution network is tiny, we assume that the phase angle difference of adjacent nodes i and node j is 0 for easy analysis. Thus, the linear voltage drop equation for line k is:

$$V_i - V_j = R_k \frac{P_{ij}}{V_i} + X_k \frac{Q_{ij}}{V_i} \tag{2}$$

where R_k and X_k are the line resistance and reactance. Equation 2 represents the node connection relationship and line parameters, describing the relationship between node voltage and power. When constructing the system, lines are numbered according to the order of the end nodes of the line. For line k, the input characteristics of the distribution network can be expressed as $X^k = \left(P_i^k, P_j^k, Q_i^k, P_j^k, V_i^k, V_i^k\right) \in \mathbb{R}_k^6$, allowing the mapping of R and X to be determined from the input X^k.

In traditional line parameter identification, the problem can be generalized as a linear regression problem or a quadratic programming problem. However, due to the fitting properties of linear regression, outliers can have significant effects on the regression, resulting in poor robustness. As the distribution network often encounters noise interference, data missing, or other situations due to the complicated operational conditions and numerous measurement devices, the performance of the linear regression model will deteriorate. Therefore, we select the deep learning method to extract the features of nodes on the premise of obtaining reconstituted measurement data samples. Deep learning techniques offer robust feature extraction capabilities, making them well-suited to handle the complexities and noise inherent in medium-voltage distribution networks.

2.2 System graph construction

Using network topology as a graph to analyze features allows for a more comprehensive utilization of structural information compared to solely relying on measurement data. Graph data $G(\mathcal{V}, \mathcal{L})$, consisting of a vertex set $\mathcal{V}$ and an edge set $\mathcal{L}$, being non-Euclidean structured, results in better classification accuracy.

In constructing the general distribution network diagram model, the bus is typically regarded as the node and the connecting line as the edge. However, graph learning focuses on node features. Therefore, for parameter identification tasks, we consider parameters as nodes of the graph and common buses between lines as edges to represent the connection relationships between lines. The feature extraction process is illustrated in the following Figure 1.

For undirected graph $G(\mathcal{V}, \mathcal{L}, A)$, V is the set of n vertices, $V_i \in V$,$\mathcal{L}$ is the set of edges in the graph, and $E_{ij} \in E$. $A \in \mathbb{R}^{n \times n}$ is the adjacency matrix, representing the topology among the nodes. Take X as the input, U is the eigenvector matrix of the normalized Laplace Matrix of the graph, g_θ is the response function of the eigenvalue. Original standard of GCN is defined as:

$$X * g_\theta = U g_\theta U^T X \tag{3}$$

Use Chebyshev polynomials $g_\theta(\Lambda) = \sum_{k=0}^{K} \theta_k T_k(\Lambda)$ to approximate and substitute it into Equation 3 to obtain Equation 4, in which $\tilde{L} = \frac{2}{\lambda_{max}} L - I_N$:

$$X * g_\theta = \sum_{k=0}^{K} \theta_k T_k(\tilde{L}) X \tag{4}$$

Then the convolution process is approximately defined as Equation 5 by using first-order Chebyshev polynomial to generate the local convolution kernel:

$$X * g_\theta = \theta\left(I_N + D^{-1/2} A D^{-1/2}\right) X = \theta\left(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}\right) X \tag{5}$$

In this formula, $\tilde{A} = A + I_N$, $D_{ii} = \sum j \tilde{A}_{ij}$. The spectral theory is applied and the output of each layer can be written as Equation 6:

$$X_k^{(l+1)} = f\left(X_k^{(l)}, A\right) = \sigma\left(\hat{D}^{-1/2} A \hat{D}^{-1/2} X_k^{(l)} W_k^{(l)}\right) \tag{6}$$

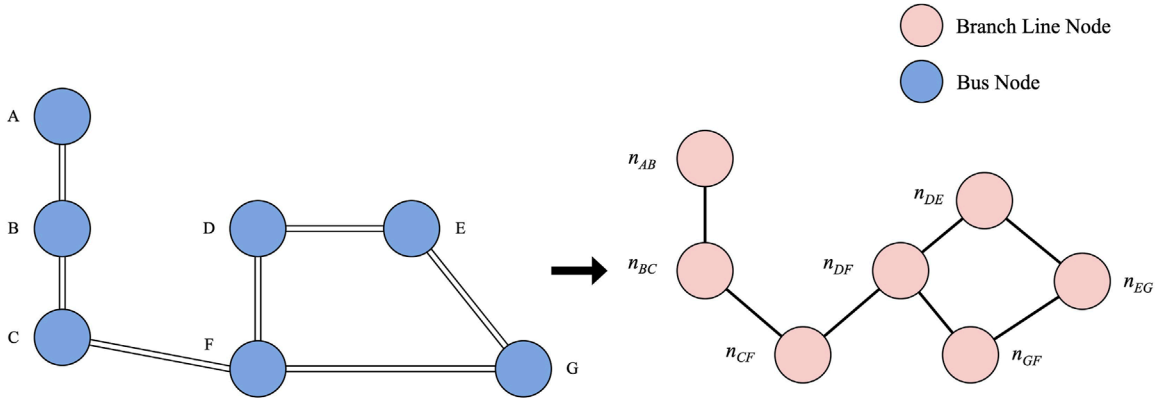


FIGURE 1
Graph extraction process.

where $X_k^{(l+1)}$ is the mapping of each layer, $l \in \{0, \ldots, L-1\}$, $X_k^{(0)}$ is the input feature X^k. $X_k^{(l)}$ represents the feature matrix of the lth layer of the model. $\sigma(\bullet)$ is the activation function, W represents the weight matrix in neural network. Since the model's input is the graph structure data X, including the adjacency matrix and corresponding attributes, the graph construction process must be completed before model training to represent the data itself and uncover the association relationships between the data.

3 GAT-MMoE model design for line parameter identification

3.1 Graph attention module design

Graph Attention Network (GAT) combines a graph neural network (GNN) with an attention mechanism, specifically tailored for processing graph-structured data by assigning different attention to neighboring nodes on a graph (Velickovic et al., 2017). It can reduce the computational cost and make it more scalable than methods that consider all neighbors equally, such as traditional Graph Convolutional Networks (GCNs). This flexibility is particularly important for large, sparse graphs. In this paper, GAT module is applied to transform the feature of each node into an inter-node attention coefficient through the graph attention layer, producing a new feature that allows for monitoring changes in neighboring nodes. Thus, information about each branch line and the distribution of impedance values between adjacent branches can be learned, improving the accuracy of parameter identification. We input $X_k^{(0)}$ and the adjacency matrix into the graph convolution layer to learn the node features and structure.

First, the voltage amplitude of nodes under a single time section is input into the graph attention network to calculate the similarity between each node and its neighbors in the distribution network. For each node i, calculate the corresponding coefficient between node j and itself:

$$e_{ij}^k = a\left(\left[W^k h_i W^k h_j\right]\right), j \in \mathcal{N}_i \tag{7}$$

where $[\cdot\cdot]$ is the concatenation operation. W^k is the learnable weight matrix for k^{th} layer of the attention mechanism, finally $a(\bullet)$ is used to map the concatenated high-dimensional features to a real number. N_i indicates the set of nodes adjacent to node i, h_i and h_j represent the feature value for node i and j respectively.

After obtaining the correlation coefficient for all the neighboring nodes of node i, the attention coefficient is normalized using softmax:

$$\alpha_{ij}^k = softmax\left(e_{ij}\right) \frac{\exp\left(\text{LeakyReLU}\left(e_{ij}^k\right)\right)}{\sum_{j \in \mathcal{N}_i} \exp\left(\text{LeakyReLU}\left(e_{ij}^k\right)\right)} \tag{8}$$

$$\text{LeakyReLU}(x) = \begin{cases} x, x > 0 \\ \beta x, x \le 0 \end{cases} \tag{9}$$

Where α_{ij}^k is the attention coefficient between the k head attention mechanism node i and the adjacent node j. According to Equations 7, 8, new node features are formed by aggregating information using the attention coefficient matrix a.

After the attention weights of all nodes are normalized, the information of nodes is extracted through the graph attention layer. For different features, different attention weights need to be assigned. If only single-layer attention is used, the same attention weights are applied to all attributes of the neighbourhood node, which will weaken the learning ability of the model. The specific calculations in each layer are shown in Equations 9–11.

In each attention layer, we use this to weigh the messages of a node's neighbours, which are the neighbour's features multiplied by the same learnable weight matrix W. We do this for each attention head and concatenate the result of the heads together:

$$h_i' = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{j \in N_i} \alpha_{ij}^k W^k h_j\right) \tag{10}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{11}$$

K is the attention head number and we choose Sigmoid as the activation function. The feature $h_i^{'}$, calculated by the multi-head attention mechanism, incorporates the contribution of the features of neighboring node j to node i, therefore having a stronger ability to

express features. The feature information is then input into the multi-task module after being learned by the GAT module to identify branch line parameters. Identification tasks for different branch parameters are input into separate expert networks, with each expert responsible for a specific subspace. Moreover, GAT possesses topological extrapolation capabilities. If the topology of the station area changes due to maintenance or other reasons, adaptive anomaly identification can be performed by inputting the new adjacency matrix after training the original adjacency matrix into GAT.

3.2 Multi-task module design

In the parameter identification work of distribution networks, multiple branches typically require parameter identification, with each branch having multiple target parameters to identify. The magnitudes of branch resistance and reactance are generally quite different, but their characteristics depend on the same factors. Moreover, branch data are strongly interrelated, and variations in the electrical variables of one transmission branch often affect the measurement data of the entire distribution network. Multi-task learning leverages the correlations between multiple tasks to optimize the performance of multi-parameter identification. In this paper, a multi-task strategy is employed to identify multiple targets simultaneously, thereby reducing computational effort. According to previous works, different branches are identified as independent tasks.

3.2.1 Model choosing and sharing policy

The Mixture of Experts (MoE) approach was initially developed and explored within the field of artificial neural networks, where experts are typically neural network models used to predict numerical values in regression or class labels in classification. To capture differences among multiple branch line parameter identification tasks, a gating network is added for each task, forming the Multi-gate Mixture of Experts (MMoE) model on the basis of MoE (Shazeer et al., 2017; Ma et al., 2018). The MMoE enhances performance by allowing multiple tasks to share a set of expert networks, while also assigning different combinations of those experts to each task. The architecture setup which contains a set of expert networks and gating networks enables better task-specific learning and reduces the risk of overfitting, especially when different tasks are related but still require some specialization. For medium-voltage distribution systems, which face constantly changing conditions like load variations and fault scenarios, traditional methods like FCNs often struggle to adapt in real-time across multiple tasks without significant re-calibration. MMoE excels in these environments by providing a more efficient and adaptive solution, improving both accuracy and reliability in parameter identification across different branches.

Multi-task learning can be divided into two mechanisms: hard parameter sharing, where different tasks share the bottom hidden layer, and soft parameter sharing. Both mechanisms have their advantages and disadvantages. In the hard sharing mechanism, parameter sharing is used for feature extraction and output, reducing the risk of overfitting. However, if task differences are large, the model results become less credible (Jacobs et al., 1991; Eigen et al., 2013). The MMoE module represents a soft parameter sharing model, using expert networks as shared substructures for parameter sharing. Each task employs a gating network to learn different combination patterns of the expert networks. Compared to the hard parameter sharing model, MMoE handles task differences more effectively and has demonstrated better performance in practice. The principle is shown in Figure 2.

The MMoE model primarily consists of two core components: Gate Net and Experts. The role of Gate Net is to establish a connection between the data and the expert model, determining which expert model should process the input sample. Experts form a relatively independent set of models, each responsible for handling a specific input subspace. First, multiple branch identification tasks are decomposed into several sub-tasks, each corresponding to a network, with an expert model trained in each subnet. We use Equations 12–14 to represent the model construction.

Let x represent the model input, for task k, the MMoE model is formulated as

$$y_k = h^k(f(x)) \tag{12}$$

$$f^k(x) = \sum_{i=1}^{n} g(x)_i f_i(x) \tag{13}$$

where n represents the number of tasks, h^k represents the specific tower network where features are fed up and analyzed. Gate networks G assign different weights to each expert, with g^k being the output of the gate network corresponding to expert network i for each subtask k. The gating network interprets the predictions made by each expert and aids in deciding which expert to trust for a given input. It takes the input pattern provided to the expert models and outputs the contribution that each expert should have in predicting the input:

$$g^k(x) = \mathrm{softmax}\left(W_{gk}x\right) \tag{14}$$

W_{gk} is a trainable matrix, $W \in \mathbb{R}^{n\times d}$. Graphs constructed in Section 2 are sparse graphs, thus fit properly as embeddings for sparse features.

For the multi-branch parameter identification task, only highly correlated experts are selected to provide accurate answers. The expert model in this paper is implemented using Multilayer Perceptrons (MLPs). Output results are obtained using pooling methods to achieve a weighted sum prediction based on expert weights. The gated model then receives data elements as input, assigns them to different expert models for inference, and outputs weights representing each expert's contribution to processing the data. The pooling system calculates a weighted sum of the classifier outputs for each class and selects the class with the highest weighted sum. To control overall sparsity, the design and parameter adjustment of the gated network are primarily relied upon when there are many learning tasks. The involvement of more expert models increases the complexity of the calculation. If the gated network activates more expert models in a single selection, model performance improves and sparsity is reduced.

3.2.2 Loss function design

In MTL, label loss is the loss in the calculation of real data labels and network prediction labels for each task. Usually, the label loss is determined by the nature of the learning task and is

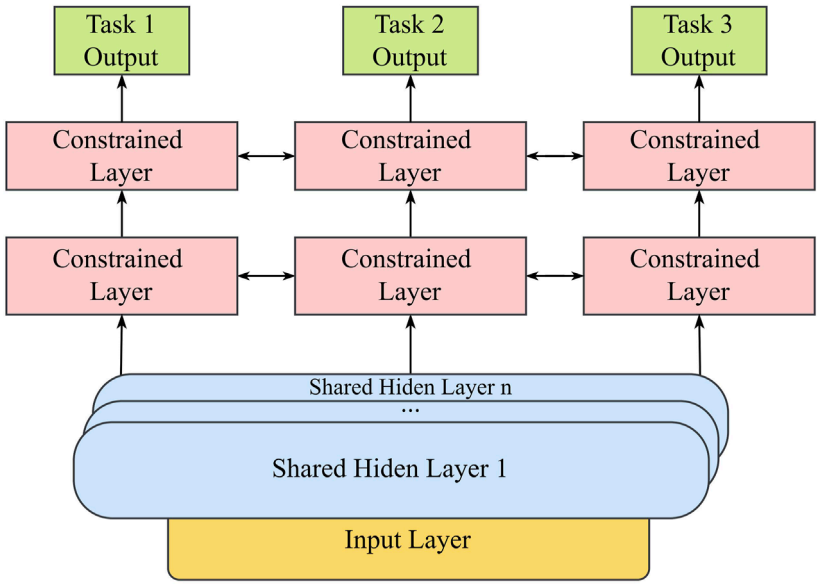


FIGURE 2
Soft-sharing mechanism.

realized through Equation 15 by weighted summation of the loss of different tasks:

$$L_{total} = \sum i\omega_i(t) * L_i \tag{15}$$

However, simply using linear weighting of multiple task losses has some significant disadvantages. Therefore, according to Cipolla et al. (2018) and Fernandez-Delgado et al. (2019), considering the distinctive contributions of different tasks to helping the final results, we use homoscedastic uncertainty as a basis for weighting losses to adjust the influence of tasks in the final loss function for optimizing the whole framework. Take two tasks, for example, the log likelihood for this output can then be written as Equations 16, 17:

$$\log p\left(y = c|f^W(x), \sigma\right) = \frac{1}{\sigma^2} f_c^W(x) - \log \sum c' exp\left(\frac{1}{\sigma^2} f_{c'}^W(x)\right) \tag{16}$$

$$p\left(y|f^W(x), \sigma\right) = Softmax\left(\frac{1}{\sigma^2} f^W(x)\right) \tag{17}$$

$f^W(x)$ is the output of a neural network with weights W on input x. σ is a positive scalar, which is learnt in the training process. Then we can attain joint loss of different tasks through Equation 18

$$\mathcal{L}(W, \sigma_1, \sigma_2) = -\log p\left(y_1, y_2|f^W(x)\right) \tag{18}$$

The multiple final loss is Equation 19:

$$L_{total} = \sum i\omega_i(t) * L_i = \sum i \frac{1}{2\sigma_i^2} L_i + \log \sigma_i \tag{19}$$

3.3 Overall framework of the proposed method

The overall framework of the proposed GAT-MMoE for distribution line parameter identification is depicted in Figure 3 As shown in Figure 3, a multi-task learning model based on an attention graph is constructed. The input of GAT-MMoE we proposed is feature matrix X and the adjacency matrix A, which means that our input features contain n nodes, each node containing six features. The final output of the whole model can be expressed as Equation 20:

$$\tilde{Y}_k = MT_k\left(X_j\right) \tag{20}$$

where MT_k represents the mapping function of the k-th branch. X_j is the j-th expert output.

To overcome the DLPI problems, the system graph is established according to the description in Section 2, with each node corresponding to a physical bus in the distribution network. The input feature of the distribution system is expressed as $\left(P_i, P_j, Q_i, Q_j, V_i, V_j\right) \in X^k$. The number of nodes is decided by the scale of distribution network, which means that our input features contain n nodes, each node contains the above six features.

The GAT module extracts the characteristics of the system and pays attention to different branch information in different subspaces by using multi-head attention mechanism. Then the different subspaces are concatenated to infuse the learnt information. The features information is then input into the multi-task module for training. The multi-task module is an MMoE-backboned MTL module, which contains multiple expert subnetworks. Gate control units are used to calculate the loss for different tasks during the training process and update the parameters related to each task based on the loss. The model's outputs are the estimated branch line impedance.

4 Case studies

In this section, the IEEE 14-node distribution network (case 1) and the IEEE 33-node distribution network (case 2) are selected

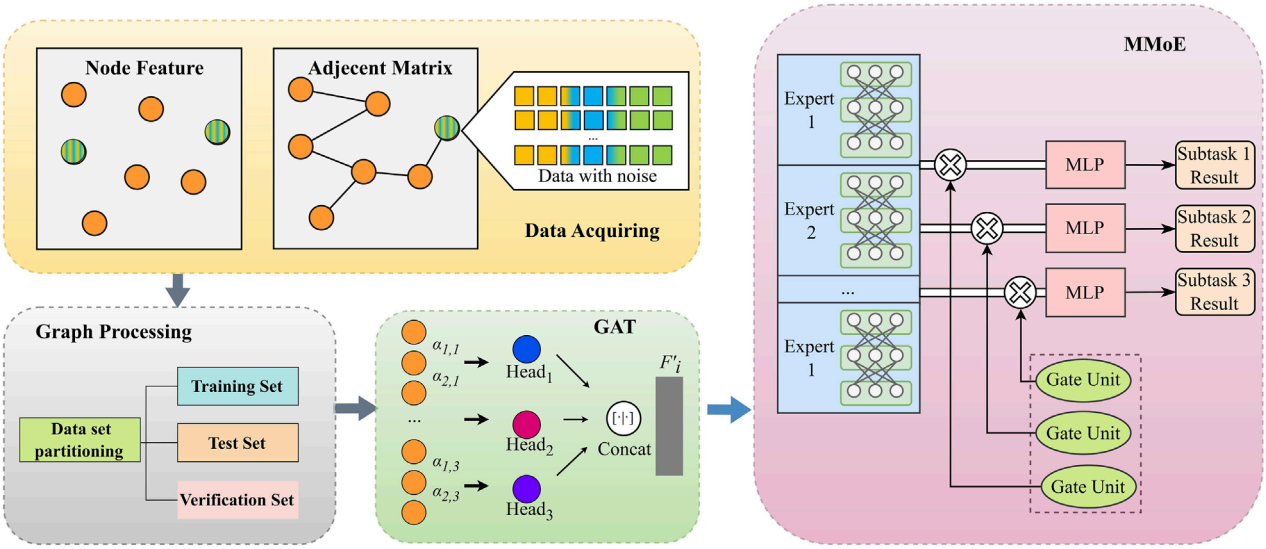


FIGURE 3
Overall framework of the proposed GAT-MMoE model.

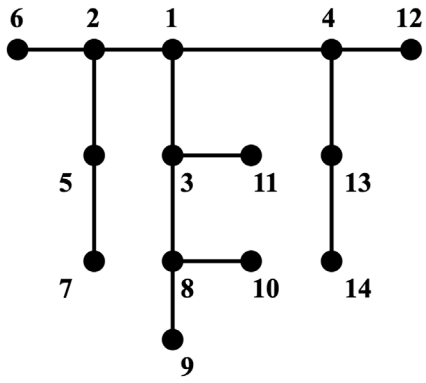


FIGURE 4
IEEE-14 distribution system.

as research objects to verify the effectiveness and robustness of the proposed method. The corresponding topological structure of the system is represented by Figures 4, 5. Experiments are performed on a computer with Intel Core i7-8700K @ 3.70 GHz CPU, and NVIDIA GeForce RTX 3060 Ti GPU. It utilizes Python3.10, Pytorch2.0.1 and pandapower2.13.1.

4.1 Dataset description

The power flow formula of power system is as follows:

$$\begin{cases} P_i = V_i \sum_{j=1}^{n} V_j \left(G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij} \right) \\ Q_i = V_i \sum_{j=1}^{n} V_j \left(G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij} \right) \end{cases} \tag{21}$$

According to Equation 21, Node voltage, active power and reactive power data are simulated by pandapower toolkit (Thurner et al., 2018). The active power injected by nodes in the load data is sampled by the Latin hypercube sampling method at $[0.8Ps, 1.2Ps]$, where Ps represents the standard active power of the distribution network. Reactive power data set is generated by active power data set and power factor. The power factor $\cos\varphi$ satisfies the uniform distribution of parameters $(0.85, 0.95)$. Reactive power is calculated by power factor through Equation 22. The formula for calculating reactive power injected by nodes is as follows:

$$Q_i(t) = \frac{P_i(t)}{\tan\left(\arccos \varphi(t)\right)} \tag{22}$$

where $\varphi(t) \in [0.85, 0.95]$, $P_i(t)$ and $Q_i(t)$ are the active power and the reactive power injected into node i at time t. $\varphi(t)$ represents the power factor of the system at time t.

In this paper, 20 types of radiative network topologies are selected, then we add Gaussian noise with σ of 0.01, 0.03, 0.5, 1, 2 to each load level, and sample each noise 6 times. A total of 69,120 sets of samples are obtained and divided, including 70% data as training set, 20% data as test set, 10% data as verification set. The hyperparameters of the model, such as the attention coefficient α, and learning rates are determined by 10% of the data set.

4.2 Evaluation index and baseline model setup

A suite of metrics is employed to manifest the performance of the model proposed in this work. Specific calculation formula as shown in Equations 23, 24:

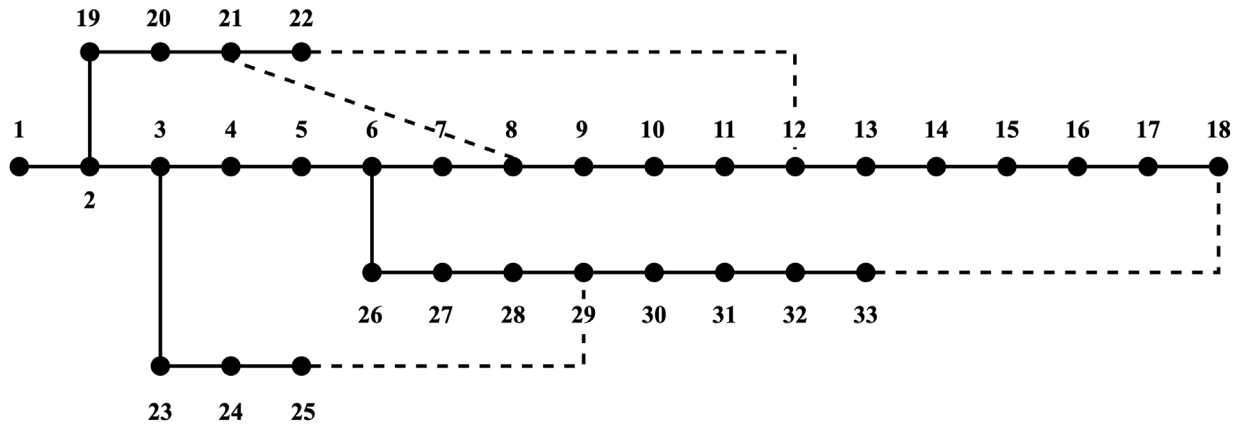


FIGURE 5
IEEE-33 distribution system.

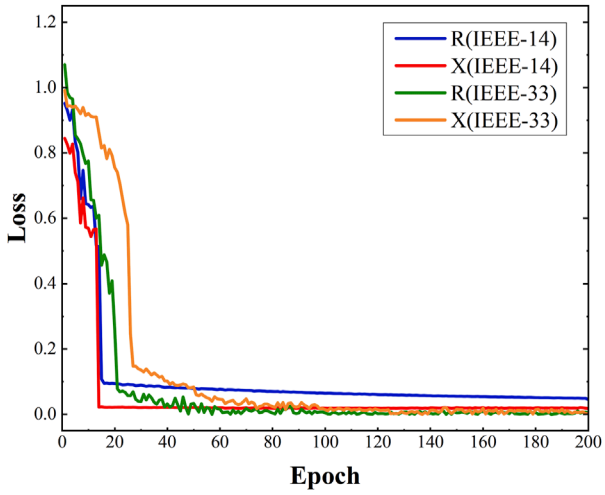


FIGURE 6
Training loss over epochs.

Root mean square error (RMSE):

$$E_{RMSE} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(y_i - \hat{y}_i)^2} \tag{23}$$

Mean absolute percentage error (MAPE):

$$E_{MAPE} = \frac{1}{K}\sum_{i=1}^{K}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{24}$$

Where y_i and $\hat{y}_i$ respectively represent the true value and the predicted value, K represents the sampling number.

In order to prove the validity of our proposed model, we adopt the following methods as baselines:

1) LR: By minimizing the sum of squares of errors, the linear regression model is used to provide coefficients that quantify the contribution of each feature to the target variable, but may face the problem of overfitting.
2) SVR: Support vector regression is a machine learning method, which adopts the idea of support vector and the Lagrange multiplier to perform regression analysis on data when doing data fitting.
3) FCN: Fully connected neural network is a type of linear neural network, which inevitably faces the problem of poor precision in dealing with nonlinear data sets and overfitting. Fully connected prediction is accomplished by flattening the input matrix.

In our implementation, the GAT module utilizes three attention heads, with a hidden representation dimensionality set to 128 and a dropout rate of 0.3. For optimization, we use a batch size of 128 and a learning rate of 0.005. The MMoE module is trained with the Adam optimizer and the learning rate is grid searched from [0.0001, 0.001, 0.01]. To prevent overfitting in both the expert and gating networks, the dropout rate is 0.2. Different weights of each task are assigned using homoscedastic uncertainty.

4.3 Results and discussion

According to Figure 6, We can find that the gradient of our proposed model decreases rapidly and converges fast after around the 20th epoch with little change in accuracy, which demonstrates the superiority of our model in deep learning-based algorithms.

From Figure 7, it can be noticed that the proposed GAT-MMoE model can effectively identify the branch line parameters. For case 1, the max relative errors lie in branch 7 and 1 respectively for R and X, reaching 3.83% and 4.35%. Table 1 shows the parameter identification errors of branch resistance and reactance, the corresponding average errors are 3.84% and 2.67% for R and X. For case 2, the relative max errors are 9.63% and 9.87%. From Table 2, average errors for R and X are 6.69% and 7.24%. The GAT-MMoE model can achieve the lowest error in most cases. The deviation comes from the Gaussian noise we add and as the resistance and reactance have different orders of magnitude, the errors are within the allowable range.

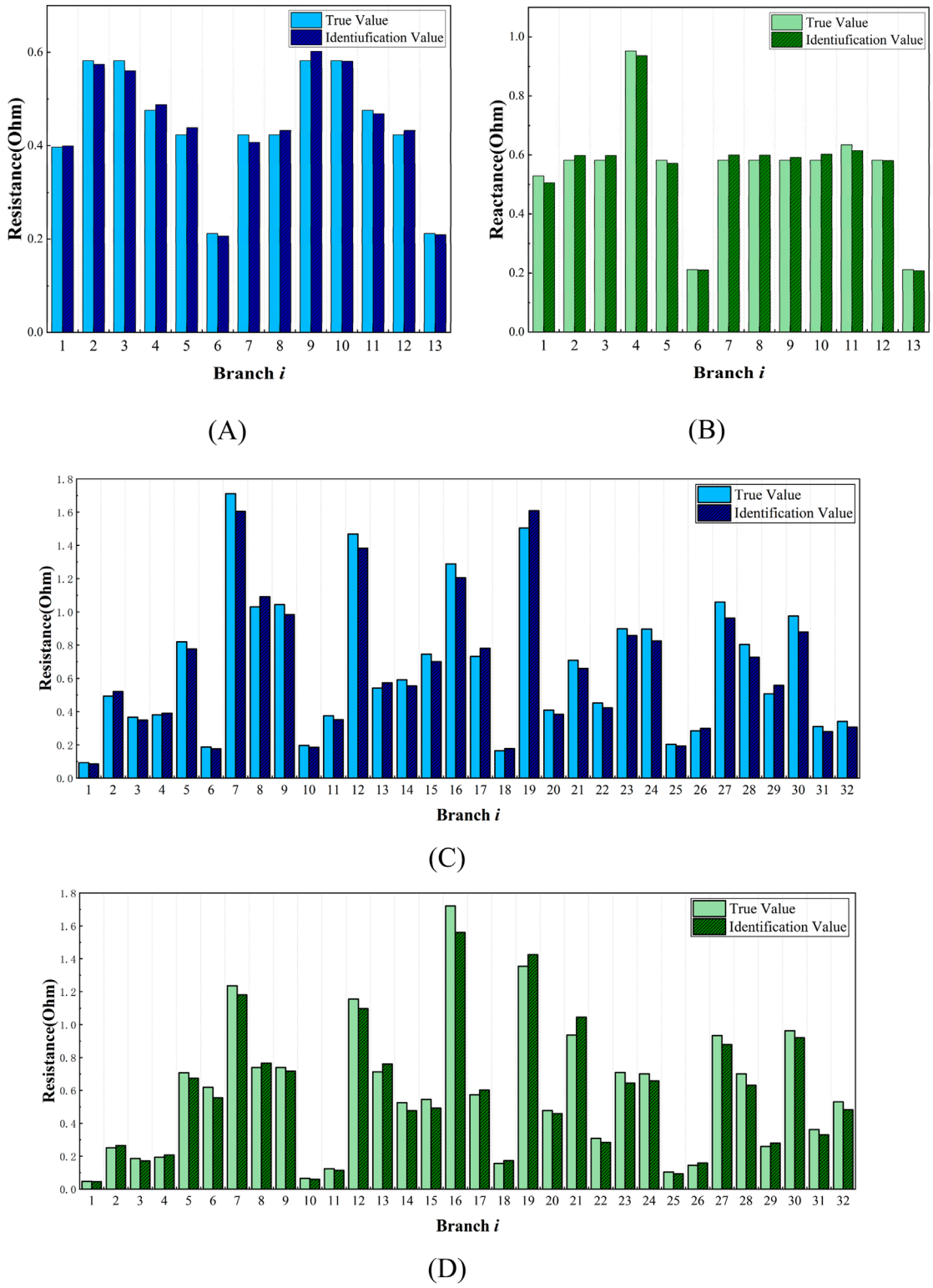


FIGURE 7
Identification results presented with true values and identification values. **(A)** line resistance of IEEE-14 system; **(B)** line reactance of IEEE-14 system; **(C)** line resistance of IEEE-33 system; **(D)** line reactance of IEEE-33 system.

Compared with the baseline model, our proposed GAT-MMoE model demonstrates higher accuracy and better robustness considering measurement error. The results in Tables 1–4 indicate that the proposed GAT-MMoE method outperforms all other models in terms of RMSE and MAPE. Network-based methods generally show superior performance compared to linear regression

TABLE 1 Identification errors in IEEE-14 system.

	LR	SVR	FCN	GAT-MMoE
Average error of R	0.3732	0.0578	0.0622	0.0384
Average error of X	0.3485	0.0563	0.0588	0.0267
Max error of R	0.1022	0.0344	0.0425	0.0218
Max error of X	0.1234	0.0437	0.0438	0.0319
Min error of R	0.4043	0.0726	0.0733	0.0083
Min error of X	0.3732	0.0794	0.0799	0.0035

TABLE 2 Identification errors in IEEE-33 system.

	LR	SVR	FCN	GAT-MMoE
Average error of R	0.3763	0.0609	0.0652	0.0569
Average error of X	0.3815	0.0593	0.0618	0.0426
Max error of R	0.4173	0.0756	0.0763	0.0638
Max error of X	0.5489	0.0824	0.0929	0.0612
Min error of R	0.1552	0.0434	0.0495	0.0168
Min error of X	0.1264	0.0467	0.0568	0.0153

TABLE 3 Identification indexes compared with different baseline models in IEEE-14 system.

R	RMSE	MAPE	X	RMSE	MAPE
LR	0.8732	0.7322	LR	0.4198	0.3485
SVR	0.1924	0.0845	SVR	0.2643	0.0967
FCN	0.2494	0.1412	FCN	0.0953	0.0876
GAT-MMoE	0.0545	0.0203	GAT-MMoE	0.0638	0.0311

TABLE 4 Identification indexes compared with different baseline models in IEEE-33 system.

R	RMSE	MAPE	X	RMSE	MAPE
LR	0.9302	0.1222	LR	0.5598	0.5697
SVR	0.2521	0.1245	SVR	0.2743	0.1267
FCN	0.6578	0.2612	FCN	0.3453	0.2384
GAT-MMoE	0.0689	0.0317	GAT-MMoE	0.0688	0.0487

and traditional machine learning methods, underscoring the value of graph attention learning in extracting high-quality features for DLP prediction. The superior performance of GAT-MMoE can be attributed to its effective utilization of related knowledge between neighboring nodes in the graph. Most baseline methods do not specifically address the issue of sparsity in their models, resulting in suboptimal performance. Our model leverages multiple data sources to construct the information network and employs a multi-task learning framework to address the specific task of predicting branch line parameters. Consequently, GAT-MMoE outperforms the selected baselines. Additionally, the model demonstrates good robustness when photovoltaic power supply is integrated into the system. Once the model training is completed, it can simultaneously predict all parameters of the power grid branch. Despite the lengthy training process, the method's robustness and accuracy compensate for this drawback. Moreover, the trained neural network model can be easily and rapidly deployed to the required locations, making it a practical solution for real-world applications.

From Tables 3, 4, different we can find that our proposed GAT-MMoE model achieves the best identification results in most cases. The indexes of multi-task learning model are better than that of single task learning model. By comparing the machine learning methods, we can find that the LR method achieves very good identification results without noise, but when the input features contain disturbance and noise, the accuracy of branch parameter identification is greatly reduced. Deep learning method like FCN fail to balance accuracy between resistance and reactance, as the loss function is only simple addition. Also, the method is more inclined to the identification result of line resistance R and ignores the branch reactance X.

The influence of distributed photovoltaic access on the proposed identification method is further explored. We incorporated multiple distributed photovoltaic (PV) systems into the distribution network, with the power data of the PV sources derived from the Desert Knowledge Australia Solar Centre (DKA Solar Center, 2024) We integrated PV1 and PV2 at nodes 7 and 12 in the IEEE 14-bus system, and at nodes 22 and 33 in the IEEE 33-bus system. Based on the sampling frequency of every 15 min, the system node data is obtained by power flow calculation. After collecting system P, Q, and V data over multiple time profiles, we input them into the model for parameter identification. Figure 8 shows that the identification error increases when a distributed power supply is present in the network. This increase is due to the changes in power flow direction caused by the integration of distributed photovoltaics. However, the identification errors remain within the acceptable range, demonstrating the robustness of the method.

For most existing distribution power system, the complexities and dynamic conditions present unique challenges and opportunities for parameter identification. Given that most existing power grid branch parameter identification methods are model-driven, resulting in low accuracy and poor reliability, our proposed model leverages a large volume of multi-source power grid operation data. It is constrained by the grid topology while integrating both local and global information. This approach allows for the comprehensive use of historical data to more accurately identify branch parameters, which can be then fed

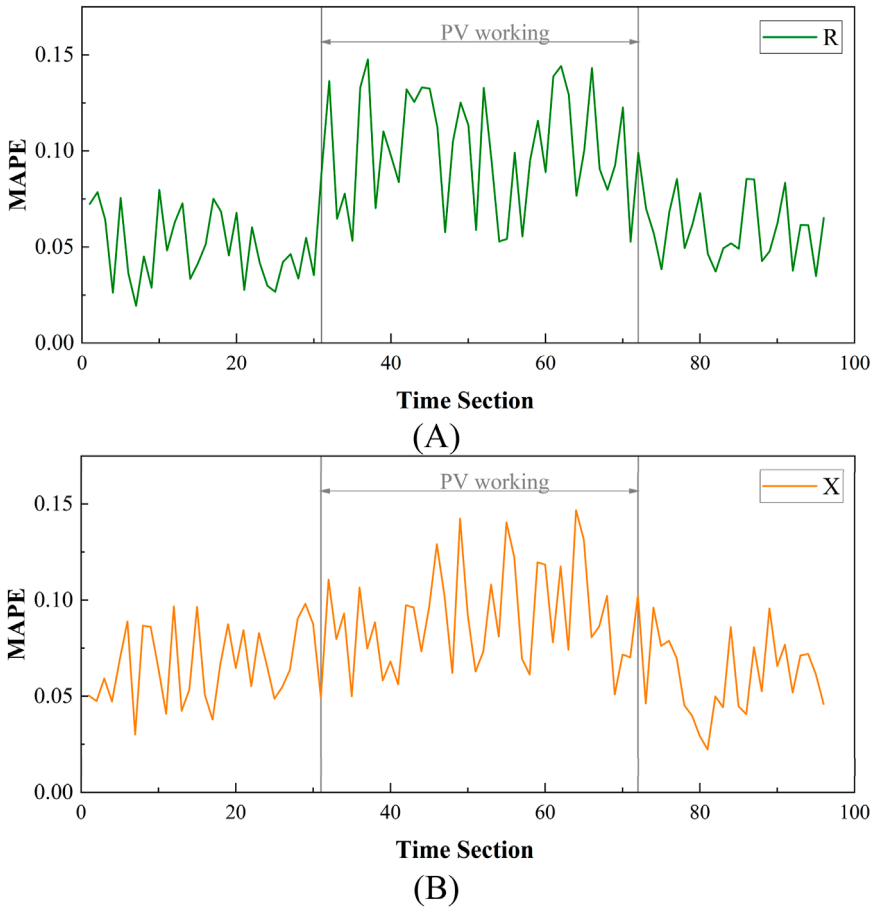


FIGURE 8
Identification results in IEEE-33 system with PV access: **(A)** line resistance MAPE of IEEE-33 system; **(B)** line reactance MAPE of IEEE-33 system.

back to the power grid dispatch center. As a result, dispatch operators gain a clearer understanding of the changing trends in branch parameters, ensuring the safe and stable operation of the power grid.

5 Conclusion

Parameter identification is crucial for distribution network scheduling and control, making it a significant research task. Current methods, which are primarily model-driven, are sensitive to data loss and noise. This paper introduces a novel line parameter identification method for medium-voltage distribution networks, considering the topology constraints of power network branches and being validated on IEEE14-M and IEEE33 systems. When there are too many layers of adjacent nodes, the global information tends to become similar, resulting in redundancy. By introducing an attention mechanism, the proposed method focuses on relatively important nodes and perform feature fusion on key branches and features. The proposed method consists of three components: graph generation, attention calculation, and multi-task prediction. The GAT module uses adaptive attention weights to flexibly model dependencies among different nodes. The MMoE algorithm addresses the coupling characteristics among multiple branches by utilizing multiple expert networks, thereby improving accuracy.

The method's effectiveness and robustness are validated through simulated grid tests, demonstrating improved results compared to traditional methods. Results show that the GAT-MMoE method achieves lower identification deviations 3.84% and 2.67% in IEEE14-M, and 5.69% and 4.26% in IEEE33 compared to the LR, SVR and FCN methods, achieving high prediction accuracy, good performance, and robustness against various types of noise.

Moreover, the GAT-MMoE method relies solely on nodal measurements of injected active power, reactive power, and voltage magnitude, streamlining the identification process without compromising accuracy. As smart grid technologies continue to evolve, data-driven deep learning approaches will play an increasingly important role in improving parameter identification in distribution networks. Future work will aim to extend this approach to a wider range of line parameters while addressing issues such as limited data availability and dynamic topology adaptation.

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

XJ: Writing–original draft, Writing–review and editing. CZ: Formal Analysis, Methodology, Writing–review and editing. QP: Validation, Visualization, Writing–review and editing. LW: Investigation, Writing–review and editing. BW: Writing–original draft. YX: Software, Writing–review and editing. KC: Resources, Writing–review and editing. LF: Supervision, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The authors received funding from the Technology Project of State Grid Co., Ltd. of Jiangsu Province (J2023018).

Acknowledgments

The authors would like to thank and acknowledge the Suzhou Power Supply Company, State Grid Jiangsu Electric Power Co., Ltd.

Conflict of interest

Authors XJ, CZ, QP, LW, BW, YX, KC, and LF were employed by Suzhou Power Supply Company, State Grid Jiangsu Electric Power Co., Ltd.

The authors declare that this study received funding from the Technology Project of State Grid Co., Ltd. of Jiangsu Province .The funder had the following involvement in the study: study design, data collection and analysis, decision to publish, preparation of the manuscript.

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

Abomazid, A. M., El-Taweel, N. A., and Farag, H. E. Z. (2022). Optimal energy management of hydrogen energy facility using integrated battery energy storage and solar photovoltaic systems. *IEEE Trans. Sustain. Energy* 13, 1457–1468. doi:10.1109/TSTE.2022.3161891

Asprou, M., and Kyriakides, E. (2018). "Identification and estimation of erroneous transmission line parameters using PMU measurements," in *2018 IEEE power and energy Society general meeting (PESGM)*, 1. doi:10.1109/PESGM.2018.8585896

Cheng, L. (2020). Equilibrium analysis of general N-population multi-strategy games for generation-side long-term bidding: An evolutionary game perspective. *J. Clea. Product.* 276, 124123. doi:10.1016/j.jclepro.2020.124123

Cheng, L., Chen, Y., and Liu, G. (2022). 2PnS-EG: a general two-population n-strategy evolutionary game for strategic long-term bidding in a deregulated market under different market clearing mechanisms. *Int. J. Electr. Power Energy Syst.* 142, 108182. doi:10.1016/j.ijepes.2022.108182

Cheng, L., Yin, L., Wang, J., Shen, T., Chen, Y., Liu, G., et al. (2021). Behavioral decision-making in power demand-side response management: a multi-population evolutionary game dynamics perspective. *Int. J. Electr. Power Energy Syst.* 276, 106743. doi:10.1016/j.ijepes.2020.106743

Cheng, L., and Yu, T. (2019). A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems. *Int. J. Energy Res.* 43, 1928–1973. doi:10.1002/er.4333

Cipolla, R., Gal, Y., and Kendall, A. (2018). "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in 2018 IEEE/CVF Conference on computer Vision and pattern Recognition, (Salt lake City, UT, USA: IEEE), 18-23 June 2018, 7482–7491. doi:10.1109/CVPR.2018.00781

DKA Solar Center (2024). *General: Desert Knowledge Australia Centre*. Alice Springs, Australia: Download Data. Alice Springs. Available at: http://dkasolarcentre.com.au/download (Accessed June 20, 2024).

Eigen, D., Ranzato, M., and Sutskever, I. (2013). Learning factored representations in a deep mixture of experts. *arXiv.Org*. Available at: https://arxiv.org/abs/1312.4314v3 (Accessed July 10, 2024). doi:10.48550/arXiv.1312.4314

Fernández-Delgado, M., Sirsat, M. S., Cernadas, E., Alawadi, S., Barro, S., and Febrero-Bande, M. (2019). An extensive experimental survey of regression methods. *Neural Netw.* 111, 11–34. doi:10.1016/j.neunet.2018.12.010

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural comput.* 3, 79–87. doi:10.1162/neco.1991.3.1.79

Li, C., Zhang, Y., Zhang, H., Wu, Q., and Terzija, V. (2018). Measurement-based transmission line parameter estimation with adaptive data selection Scheme. *IEEE Trans. Smart Grid* 9, 5764–5773. doi:10.1109/TSG.2017.2696619

Li, H., Weng, Y., Vittal, V., and Blasch, E. (2024). Distribution grid topology and parameter estimation using deep-shallow neural network with physical consistency. *IEEE Trans. Smart Grid* 15, 655–666. doi:10.1109/TSG.2023.3278702

Li, X., Wang, S., and Lu, Z. (2022). Reverse identification method of line parameters in distribution network with multi-T nodes based on partial measurement data. *Electr. Pow. Syst. Res.* 204, 107691. doi:10.1016/j.epsr.2021.107691

Li, Y., Zhang, H., Liang, X., and Huang, B. (2019). Event-triggered-based distributed cooperative energy management for multienergy systems. *IEEE Trans. Ind. Inf.* 15, 2008–2022. doi:10.1109/TII.2018.2862436

Lin, Y., and Abur, A. (2018). Strategic use of synchronized phasor measurements to improve network parameter error detection. *IEEE Trans. Smart Grid* 9, 5281–5290. doi:10.1109/TSG.2017.2686095

Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., and Chi, E. H. (2018). "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts," in Proceedings of the 24th ACM SIGKDD International Conference on knowledge Discovery & data mining, 1930–1939. doi:10.1145/3219819.3220007

Pegoraro, P. A., Brady, K., Castello, P., Muscas, C., and von Meier, A. (2019a). Compensation of systematic measurement errors in a PMU-based monitoring system for electric distribution grids. *IEEE Trans. Instrum. Meas.* 68, 3871–3882. doi:10.1109/TIM.2019.2908703

Pegoraro, P. A., Brady, K., Castello, P., Muscas, C., and von Meier, A. (2019b). Line impedance estimation based on synchrophasor measurements for power distribution systems. *IEEE Trans. Instrum. Meas.* 68, 1002–1013. doi:10.1109/TIM.2018.2861058

Pegoraro, P. A., Castello, P., Muscas, C., Brady, K., and von Meier, A. (2017). "Handling instrument transformers and PMU errors for the estimation of line parameters in distribution grids," in *2017 IEEE international workshop on applied measurements for power systems (AMPS)*, 1–6. doi:10.1109/AMPS.2017.8078339

Pegoraro, P. A., Sitzia, C., Solinas, A. V., and Sulis, S. (2022). PMU-based estimation of systematic measurement errors, line parameters, and tap changer ratios in three-phase power systems. *IEEE Trans. Instrum. Meas.* 71, 1–12. doi:10.1109/TIM.2022.3165247

Puddu, R., Brady, K., Muscas, C., Pegoraro, P. A., and Von Meier, A. (2018). "PMU-based technique for the estimation of line parameters in three-phase electric distribution grids," in *2018 IEEE 9th international workshop on applied measurements for power systems (AMPS)*, 1–5. doi:10.1109/AMPS.2018.8494886

Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., et al. (2017). Outrageously large neural networks: the sparsely-gated mixture-of-experts layer. doi:10.48550/arXiv.1701.06538

Shi, X., Qiu, R., He, X., Ling, Z., Yang, H., and Chu, L. (2020). Early anomaly detection and localisation in distribution network: a data-driven approach. *IET Gener. Transm. Distrib.* 14, 3814–3825. doi:10.1049/iet-gtd.2019.1790

Shi, Z. ., Xu, Q., Liu, Y., Wu, C., and Yang, Y. (2024). Line parameter, topology and phase estimation in three-phase distribution networks with non-μPMUs (2024). *Int. J. Electr. Power Energy Syst.* 155, 109658. doi:10.1016/j.ijepes.2023.109658

Singh, R. S., Cobben, S., Gibescu, M., van den Brom, H., Colangelo, D., and Rietveld, G. (2018). "Medium voltage line parameter estimation using synchrophasor data: a step towards dynamic line rating," in *2018 IEEE power and energy society general meeting (PESGM)*, 1–5. doi:10.1109/PESGM.2018.8586111

Sun, J., Xia, M., and Chen, Q. (2019). A classification identification method based on phasor measurement for distribution line parameter identification under insufficient measurements conditions. *IEEE Access* 7, 158732–158743. doi:10.1109/ACCESS.2019.2950461

Thurner, L., Scheidler, A., Schäfer, F., Menke, J.-H., Dollichon, J., Meier, F., et al. (2018). Pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Trans. Power Syst.* 33, 6510–6521. doi:10.1109/TPWRS.2018.2829021

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio', P., and Bengio, Y. (2017). Graph attention networks. *ArXiv*. doi:10.48550/arXiv.1710.10903

Wang, W., and Yu, N. (2022). Estimate three-phase distribution line parameters with physics-informed graphical learning method. *IEEE Trans. Power Syst.* 37, 3577–3591. doi:10.1109/TPWRS.2021.3134952

Wang, Y., Xu, X., and Xue, H. (2016). Method to measure the unbalance of the multiple-circuit transmission lines on the same tower and its applications. *IET Gener. Transm. Distrib.* 10, 2050–2057. doi:10.1049/iet-gtd.2015.0979

Wang, Z., Xia, M., Lu, M., Pan, L., and Liu, J. (2022). Parameter identification in power transmission systems based on graph convolution network. *IEEE Trans. Power Del.* 37, 3155–3163. doi:10.1109/TPWRD.2021.3124528

Xia, M., Wang, Z., Lu, M., and Pan, L. (2022). MFAGCN: a new framework for identifying power grid branch parameters. *Electr. Pow. Syst. Res.* 207, 107855. doi:10.1016/j.epsr.2022.107855

Xiao, M., Xie, W., Fang, C., Wang, S., Li, Y., Liu, S., et al. (2021). Distribution line parameter estimation driven by probabilistic data fusion of D-PMU and AMI. *IET Gener. Transm. Distrib.* 20, 2883–2892. doi:10.1049/gtd2.12224

Yang, L., Li, X., Sun, M., and Sun, C. (2023). Hybrid policy-based reinforcement learning of adaptive energy management for the energy transmission-constrained island group. *IEEE Trans. Ind. Inf.* 19, 10751–10762. doi:10.1109/TII.2023.3241682

Yang, N.-C., Huang, R., and Guo, M.-F. (2022). Distribution feeder parameter estimation without synchronized phasor measurement by using radial basis function neural networks and multi-run optimization method. *IEEE Access* 10, 2869–2879. doi:10.1109/ACCESS.2021.3140123

Yu, J., Weng, Y., and Rajagopal, R. (2018). PaToPa: a data-driven parameter and topology joint estimation framework in distribution grids. *IEEE Trans. Power Syst.* 33, 4335–4347. doi:10.1109/TPWRS.2017.2778194

Yu, J., Weng, Y., and Rajagopal, R. (2019). PaToPaEM: a data-driven parameter and topology joint estimation framework for time-varying system in distribution grids. *IEEE Trans. Power Syst.* 34, 1682–1692. doi:10.1109/TPWRS.2018.2888619

Zhang, Y., Wang, J., and Li, Z. (2020). Interval state estimation with uncertainty of distributed generation and line parameters in unbalanced distribution systems. *IEEE Trans. Power Syst.* 35, 762–772. doi:10.1109/TPWRS.2019.2926445

Zhu, J., and Abur, A. (2010). Improvements in network parameter error identification via synchronized phasors. *IEEE Trans. Power Syst.* 25, 44–50. doi:10.1109/TPWRS.2009.2030274

Zou, G., Xia, M., Zhang, L., Lei, Z., Peng, Z., and Liu, J. (2024). Intelligent identification of power grid parameters based on dynamic weighting. *Eng. Appl. Artif. Intell.* 135, 108822. doi:10.1016/j.engappai.2024.108822

frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 27 November 2024
DOI 10.3389/fenrg.2024.1422728

Check for updates

OPEN ACCESS

EDITED BY
Yang Yang,
Nanjing University of Posts and Telecommunications, China

REVIEWED BY
Linfei Yin,
Guangxi University, China
Yushuai Li,
Aalborg University, Denmark

*CORRESPONDENCE
ShouSheng Zhao,
zhao_shou_sheng@163.com

RECEIVED 24 April 2024
ACCEPTED 09 July 2024
PUBLISHED 27 November 2024

CITATION
Zhao S, Yang X, Li K, Li X, Qi W and Huang X (2024), Photovoltaic output prediction based on VMD disturbance feature extraction and WaveNet.
Front. Energy Res. 12:1422728.
doi: 10.3389/fenrg.2024.1422728

COPYRIGHT
© 2024 Zhao, Yang, Li, Li, Qi and Huang. 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.

Photovoltaic output prediction based on VMD disturbance feature extraction and WaveNet

ShouSheng Zhao*, Xiaofeng Yang, Kangyi Li, Xijuan Li, Weiwen Qi and Xingxing Huang

State Grid Zhejiang Electric Power Co., Ltd., Shaoxing Power Supply Company, Shaoxing, Zhejiang, China

Traditional photovoltaic (PV) forecasting methods often overlook the impact of the correlation between different power fluctuations and weather factors on short-term forecasting accuracy. To address this, this paper proposes a PV output forecasting method based on Variational Mode Decomposition (VMD) disturbance feature extraction and the WaveNet model. First, to extract different feature variations of the output and enhance the model's ability to capture PV power fluctuation details, VMD is used to decompose the PV output time series, obtaining IMFs modes representing output disturbances and quasi-clear sky IMF modes. Then, to reveal power changes, especially the underlying patterns of disturbances and their relationship with weather factors, K-means clustering is applied to the IMF modes representing output disturbances, clustering the disturbance IMFs into different power change feature clusters. This is combined with Spearman correlation analysis of weather factors and the construction of an experimental dataset. Finally, to enhance the model's learning ability and improve short-term output forecasting accuracy, the WaveNet model is employed during the forecasting phase. Separate WaveNet models are constructed and trained with the corresponding datasets, and the total PV output forecast is obtained by superimposing the predictions of different IMF modes. Experimental results are compared with traditional methods, demonstrating a significant improvement in forecasting accuracy, with a Mean Absolute Percentage Error (MAPE) error of 6.94%, highlighting the effectiveness of our method and providing strong technical support for the refined management and intelligent forecasting of PV energy.

KEYWORDS

photovoltaic output prediction, VMD, K-means, spearman, WaveNet

1 Introduction

As a clean and renewable energy source, PV power generation plays an increasingly important role in the global energy transition and the development of renewable energy. Traditional PV output prediction methods mainly rely on artificial intelligence methods and numerical weather forecasting to predict future PV output. However, changes in lighting conditions often have a significant impact on the output time series, especially in cases of abrupt changes in short-term lighting due to weather variations, resulting in large fluctuations in PV power. In recent years, researchers have actively explored various methods to predict PV output accurately and anticipate power fluctuations. With the advancement of computer technology, data-driven artificial intelligence algorithms have been widely applied in PV power prediction (Miao et al., 2023; Dong et al., 2024).

For short-term PV forecasting, literature (Dong et al., 2023) introduced a method based on the Improved Grey Wolf Optimization (IGWO) algorithm and Spiking Neural Network (SNN) for short-term PV output prediction. In the field of ultra-short-term PV power forecasting based on deep learning, Raiker proposed an ultra-short-term PV power forecasting model based on optimal frequency-domain decomposition and deep learning. The model uses convolutional neural networks to predict the low-frequency and high-frequency components separately, and then reconstructs the final prediction result through addition, significantly improving prediction accuracy and time efficiency (Raiker et al., 2021). Addressing the issue of data quality dependence in PV power model prediction, another literature (Wang et al., 2022) proposed a combination prediction method for ultra-short-term PV power generation by integrating Singular Spectrum Analysis and Local Emotion Reconstruction Neural Network. Recognizing the tendency of traditional Extreme Learning Machines to fall into local optimums and the characteristics of environmental changes causing PV output fluctuations, literature (Cheng et al., 2023) constructed a PV output short-term prediction model by employing an Adaptive Noise Complete Ensemble Empirical Mode Decomposition (CEEMDAN) algorithm combined with chimp optimization algorithm (Ceyhun and Hakan, 2021; Leiming et al., 2023) to optimize the Extreme Learning Machine neural network (Muqaddas et al., 2022). Utilizing the CEEMDAN algorithm, critical environmental factors affecting PV output power are decomposed to obtain local features of data signals at different time scales, reducing the non-stationarity of environmental factor sequences. Then, each decomposed subsequence and historical PV data sequence are used as inputs to the Extreme Learning Machine prediction model optimized by the chimp optimization algorithm for prediction. To address the incompleteness in considering the volatility of PV output and meteorological factors, literature (Bian and Sun, 2021) proposed an improved Typical Meteorological Year (TMY) method to generate representative meteorological data. This method constructs a dataset by selecting specific monthly data that best represent long-term average meteorological characteristics. Specifically, it uses metrics like Root Mean Square Error (RMSE) and correlation coefficients to choose the months with the smallest errors and highest correlations, forming a complete TMY dataset. This dataset, combined with the Generalized Regression Neural Network (GRNN) (Zhuang et al., 2019), is used for PV power prediction, thereby improving the accuracy and reliability of the predictions. Another literature (Jin et al., 2024) utilized clustering algorithms to cluster raw data and implemented PV power prediction using Long Short-Term Memory (LSTM) neural networks. They also employed an improved Sparrow Search Algorithm for neural network hyperparameter optimization, achieving optimization for different power feature scenarios. In enhancing the accuracy of PV output interval prediction, literature (Zhang C. et al., 2023) introduced a PV output interval prediction model based on Improved Ensemble Empirical Mode Decomposition and Quasi-Affine Transformation optimized Bidirectional Long Short-Term Memory neural networks (Zhu et al., 2020; Zhang et al., 2024). Additionally, literature (Wu et al., 2023) proposed a support vector machine PV power interval short-term prediction model based on Ensemble Empirical Mode Decomposition and Chaos Ant-Lion Algorithm. In terms of spatial correlation analysis, M. Zhang proposed a short-term solar power forecasting method based on an optimal graph structure that considers surrounding spatio-temporal correlations. This method improves forecasting performance by utilizing spatial information from neighboring photovoltaic stations combined with a graph convolutional network (Zhang M. et al., 2023). In terms of hybrid forecasting methods, X. Zhang proposed a new digital twin (DT) supported PV power prediction framework. This framework ensures reliable data transmission and leverages the advantages of both digital physical models and neural network models, thereby improving prediction accuracy (Zhang X. et al., 2023). In the field of deep learning networks based on satellite cloud images, Cheng proposed a graph learning framework. This framework generates directed graphs by simulating cloud movements and applies a spatio-temporal graph neural network, effectively improving the accuracy of photovoltaic power prediction while reducing image input redundancy (Cheng et al., 2022).

Although the aforementioned methods have achieved promising prediction results, they still have some limitations. Firstly, these methods may not fully capture all the factors affecting PV output when dealing with complex weather conditions and sudden environmental changes. Therefore, it is necessary to enhance the ability to identify weather changes, environmental conditions, and internal noise to more accurately capture the root causes of PV output fluctuations. Secondly, these methods have not deeply studied the characteristics and variation patterns of various output fluctuations. These methods also have not thoroughly considered the correlations between various disturbances and weather factors, resulting in a need for improved prediction accuracy under changing weather conditions. Therefore, there is a need to establish more comprehensive predictive models that consider various factors' influences to enhance the understanding and predictive ability of PV output fluctuations.

To address the low prediction accuracy of existing PV power prediction techniques and the weak correlation between meteorological factors and power fluctuations, this paper proposes a PV output prediction method based on VMD and WaveNet. Firstly, to extract different feature variations of the output, VMD (Meng et al., 2023; Parri et al., 2024; Wang and Ma, 2024; Yagang et al., 2024) is utilized to decompose the PV output time series, obtaining Intrinsic Mode Functions (IMFs) modes representing output disturbances and quasi-clear sky IMF modes. Subsequently, K-means clustering is applied to the IMFs modes representing output disturbances to cluster the disturbance IMFs into different power change feature clusters (Sleiman and Su, 2024). Spearman correlation analysis is then conducted on different feature clusters combined with weather factors to construct an experimental dataset. Lastly, to enhance the model's learning ability, a WaveNet model (Pramono et al., 2019; Deng et al., 2022; Wang H. et al., 2023; Wang Y. et al., 2023) is employed in the prediction phase. WaveNet is selected due to its superior capabilities in handling time-series data. It effectively processes long-term dependencies through its dilated convolution structure, captures multi-scale temporal features with its deep convolutional layers, and maintains robustness and stability with residual connections. Moreover, WaveNet's ability to model non-linear relationships makes it particularly suited for PV output

prediction, which involves complex interactions between various factors. Based on the input of the corresponding feature IMF time series and combined with relevant meteorological data, WaveNet models are separately constructed for training and prediction. The predicted results of different IMF modes are then superimposed to obtain the total PV output prediction. The effectiveness and accuracy of the proposed method are validated using historical data from a PV station in Zhejiang, China.

2 Power feature extraction

2.1 VMD power feature decomposition

VMD is a method used for decomposing signals and extracting different frequency features. In this paper, VMD is employed to decompose the PV output time series into different IMFs modes, which reflect varying patterns at different time scales. VMD decomposes the original time series data into multiple IMFs with different frequency characteristics, thereby better representing the feature variations of the output. This facilitates the analysis of quasi-clear sky and output disturbance characteristics, clustering them into feature clusters, laying the foundation for subsequent PV output prediction.

During the VMD decomposition, the sum of the expected bandwidths for each power IMF mode is minimized, with the constraint that the sum of all decomposition modes equals the original output feature signal sequence. The constrained variational problem is formulated as follows (Equation 1):

$$\begin{cases} \min\limits_{\{u_k\},\{w_k\}} \left\{ \sum\limits_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{pt} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ s.t. \quad \sum\limits_k u_k = f \end{cases} \tag{1}$$

In the equation, f represents the original time series, $\delta(t)$ is the Dirac distribution function, $\{u_k\}=\{u_1,\ldots,u_k\}$ and $\{w_k\}:=\{w_1,\ldots,w_k\}$ are shorthand symbols for all modes and their corresponding center frequencies, respectively. $e^{-j\omega kt}$ represents the exponential term at the respective center frequency w_k for mode u_k.

By introducing Lagrange multipliers to transform the inequality constraint into an equality constraint, and then solving the above equation, the solution formula for mode u_k can be obtained as follows (Equation 2):

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2} \tag{2}$$

In the equation, α is a quadratic penalty factor used to balance the trade-off between the objective function and the degree of violation of the constraint. By penalizing the constraint, the algorithm is encouraged to converge towards solutions that satisfy the constraint. λ is the Lagrange multiplier operator.

The formula for the center frequency w_k is (Equation 3):

$$w_k^{n+1} = \frac{\int_0^\infty w|\hat{u}_k(w)|^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \tag{3}$$

In PV output prediction, quasi-clear sky IMFs and output disturbance IMFs have different physical meanings and predictive patterns. Quasi-clear sky IMFs mainly reflect the basic characteristics of PV output under clear sky conditions, while output disturbance IMFs reflect the influence of other factors (such as cloud cover, temperature changes, etc.) on PV output. Separating quasi-clear sky and output disturbance IMFs can allow the prediction model to capture different types of variations more finely, thereby improving prediction accuracy.

2.2 Disturbance IMF clustering

After obtaining the quasi-clear sky IMF and various disturbance IMFs, clustering operations are performed on the disturbance IMFs. K-means is a clustering algorithm (Sleiman and Su, 2024) that partitions data points into different clusters based on their feature similarity. In this paper, K-means is used to cluster the IMFs modes representing PV output disturbances, grouping these modes into different clusters of power change features to better understand and describe the operating characteristics of PV power generation systems. The specific algorithm process is as follows:

1) Initialize the centroids for the disturbance IMF clusters and select the number of clusters, K.
2) Assign samples D and calculate the Euclidean distance between each sample point and the cluster centroids C_i. Find the optimal distance and assign the sample points to the feature clusters corresponding to C_i (Equation 4):

$$d(x, C_i) = \sqrt{\sum_{j=1}^{m} (x_j - C_{ij})^2} \tag{4}$$

In the equation, C_i represents the ith cluster centroid, m is the dimensionality of the data objects, and C_{ij} denote the jth attribute values of x and C_i, respectively.

3) Update the cluster centroids by computing the mean and squared error of all points in each cluster. Update the centroids and repeat step 2). The calculation formula is as follows (Equation 5):

$$\sum_{i=1}^{k} \sum_{x \in C_i} |d(x, C_i)|^2 \tag{5}$$

4) When the cluster centroids no longer change or reach the maximum number of iterations, stop the loop, update the clustering results, and calculate evaluation metrics. For different numbers of clusters K. For the disturbance IMFs, this paper calculates the Davies-Bouldin index (DBI) and the silhouette coefficient index (SC) to select the optimal index and its corresponding number of clusters KK as well as the corresponding clustering situation as the clustering result.

2.3 Spearman correlation analysis

Spearman correlation is a non-parametric method used to measure the monotonic relationship between two variables. It is robust and not influenced by outliers, making it suitable for various types of data analysis, particularly effective in detecting nonlinear relationships.

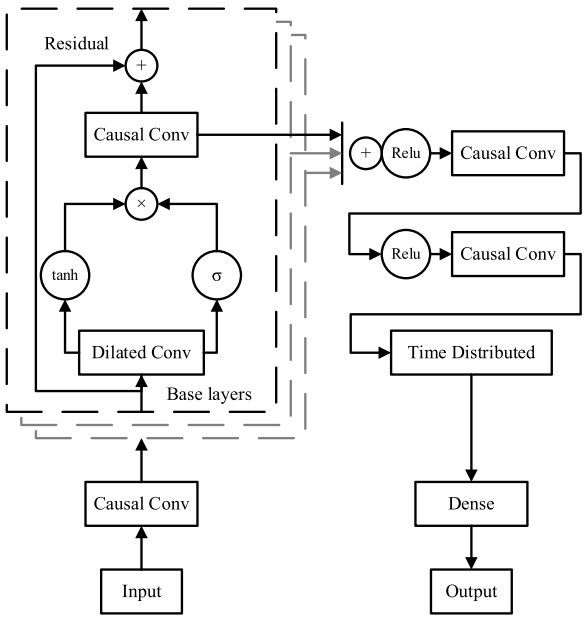


FIGURE 1
The structure diagram of the WaveNet model.

By employing Spearman correlation analysis, we can assess the degree of association between different IMF feature clusters and weather features. This helps identify which weather factors have a significant impact on different PV output features, aiding in the selection of the most relevant features to guide model construction and prediction processes. The formula for calculating the Spearman correlation coefficient is as follows (Equation 6):

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{6}$$

In the equation, ρ represents the Spearman correlation coefficient, d_i denotes the difference between the ranks of each corresponding pair (i.e., the difference between the rankings of IMF variables and weather feature variables), n is the number of data pairs.

3 WaveNet model

The WaveNet model, based on convolutional neural networks (CNNs) with different structures, is essentially a probabilistic autoregressive model for time series data. It has shown good performance in audio analysis applications, utilizing its strong capability in handling time-series features to improve short-term forecasting effects. In PV prediction, the WaveNet model can effectively capture the temporal features and nonlinear relationships in PV output data, thereby enhancing prediction accuracy and generalization capability.

The basic module of WaveNet mainly consists of dilated convolutional structures, residual connections, and gating unit structures, as shown in Figure 1. The input layer uses causal convolutions to preserve the positional information of the model's time-series feature input, preventing the model from seeing the entire temporal information at once during learning. Its convolutional structure includes causal convolutions and dilated convolutions, connecting convolutional layers with different dilation rates to obtain an ultra-long receptive field, extracting time-series features of different lengths of PV output changes, analyzed by the causal convolutional layers. The computation of the model's gating units is shown in Equation 7, and the calculation formula for the input-output of a single filter in the dilated convolutional layer is shown in Equation 8.

$$\boldsymbol{z} = \tanh\left(W_f * \boldsymbol{x}\right) \odot \sigma\left(W_g * \boldsymbol{x}\right) \tag{7}$$

In the equation, x represents the input time series to the gating unit, W_f and W_g are the corresponding weight parameters for the gating mechanism input.

$$\boldsymbol{y} = \boldsymbol{x}(t) * f(t) = \sum_{n=0}^{K-1} f(t)\boldsymbol{x}(t - dn) \tag{8}$$

In the equation, y represents the output of a single filter in the dilated convolutional layer, $x(t)$x(t) is the time series input to the dilated convolutional layer, $f(t)$ is the filter with a kernel size of k, d is the convolutional dilation rate, and n is the convolutional kernel index.

The complete WaveNet model is formed by stacking multiple basic dilated convolutional layers, which process very long time series data through stacking multiple identical parameterized basic structures. The lower layers of the convolutional structure learn short-term patterns, while long-term patterns are learned by higher layers of convolutional layers. Additionally, a residual network structure is employed to address the problem of gradient vanishing and exploding during training caused by excessive model depth. The model's output is fused using skip connections, which combine the feature quantities extracted at different convolutional layer levels, and the final prediction result is outputted through multiple causal convolutions.

4 PV output prediction based on WaveNet

Based on the above methods, this paper proposes a PV output prediction method based on VMD disturbance feature extraction and WaveNet model. The structure of the prediction model is shown in Figure 2.

First, to extract the different feature changes in the output, this paper adopts VMD to decompose the PV output time series, obtaining the IMFs modes representing output disturbances and quasi-clear sky IMF modes. The input for this stage is the historical PV output data, and the output is the decomposed IMFs modes.

The PV power generation varies rhythmically with the alternation of day and night. During the day, when the sunlight intensity is high, the power generation increases. Conversely, during the night, when the sunlight diminishes, the power generation decreases. This rhythmic variation is an inherent characteristic of PV power generation induced by the rotation and revolution of the Earth. However, sudden changes in meteorological conditions can affect the output power of the PV system, causing irregular

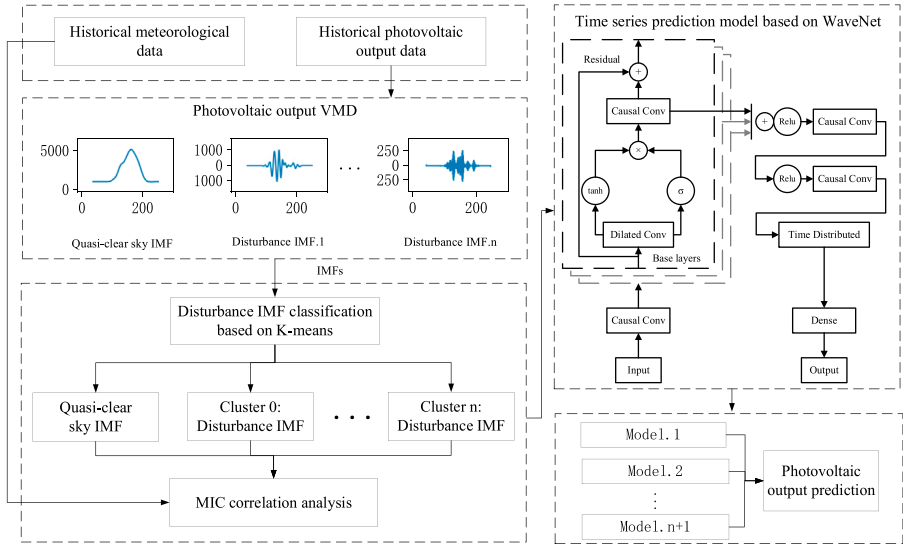


FIGURE 2
Flowchart of photovoltaic output prediction based on VMD disturbance feature extraction and WaveNet neural network.

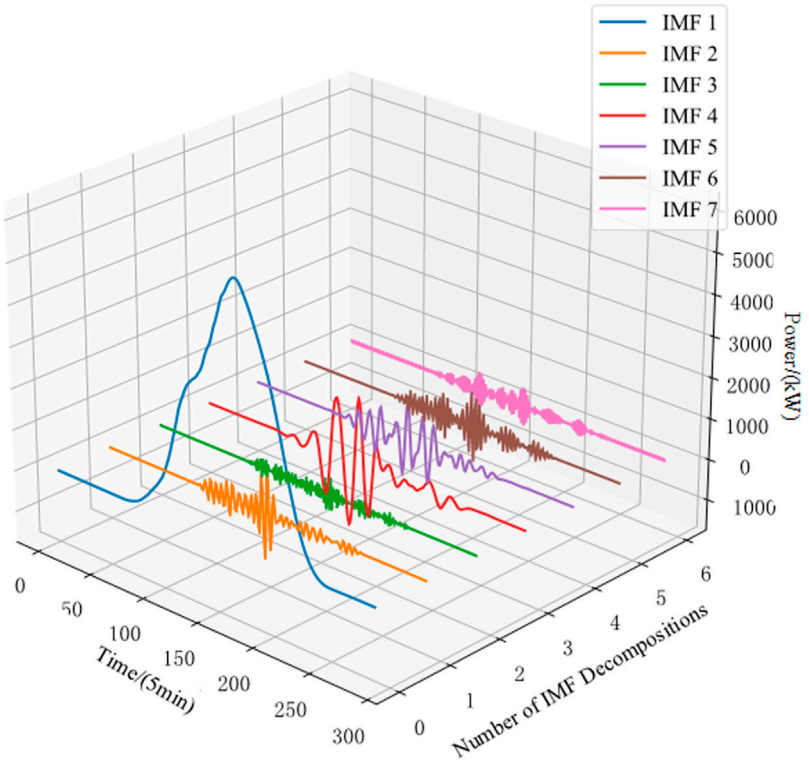


FIGURE 3
Example diagram of VMD feature decomposition.

fluctuations. By using VMD to extract quasi-clear sky curves and disturbance curves, as shown in Figure 3, this paper reveals that the quasi-clear sky curve reflects the regular changes in the output of the PV system, while the disturbance curve reflects the irregular fluctuations caused by changes in meteorological conditions. By decomposing and analyzing regular and irregular variations, a better understanding of the characteristics and variation patterns of the PV system's output can be achieved.

Next, the IMFs modes representing the output disturbance are subjected to K-means clustering to cluster the disturbance IMFs according to their power variation characteristics and plot their cluster centroids. The input for this stage is the IMFs modes representing output disturbances, and the output is the clustered disturbance IMFs and their centroids. This clustering method helps capture different types of power fluctuation patterns in the experimental dataset. By analyzing these clustered feature clusters, we can better understand the impact of different types of disturbances on PV output, providing more information and features for subsequent prediction models.

Establishing the experimental dataset is a crucial step in PV output prediction research. The input for this stage is the clustered disturbance IMFs and historical meteorological data, and the output is the experimental dataset for model training and validation. By clustering the disturbance IMFs into different power variation feature clusters and conducting Spearman correlation analysis based on their cluster centroids, optimal weather features for each feature cluster are selected, facilitating the construction of an experimental dataset for model training and validation. In the experimental dataset, each feature cluster represents a type of power fluctuation pattern and contains relevant weather data samples for that pattern. Constructing the dataset in this way helps train the model to better adapt to different types of power variation scenarios.

Finally, to further enhance the model's learning capability, the WaveNet recurrent neural network is employed in the prediction stage. The input for this stage is the experimental dataset consisting of feature IMF time series data and relevant meteorological data, and the output is the predicted PV output for each IMF mode. By combining the corresponding feature IMF time series data with relevant meteorological data, a WaveNet model is constructed for training and prediction. WaveNet is a convolutional neural network structure composed of a series of convolutional layers, each containing multiple convolutional kernels. These kernels have gradually expanding receptive fields, allowing the network to capture rich information at different time scales. During the prediction process, the corresponding feature IMF time series data is combined with relevant meteorological data to train and predict using the WaveNet model. WaveNet can effectively handle time series data and extract important feature information, aiding in better understanding the spatiotemporal structure and related properties of the data. WaveNet itself has strong nonlinear modeling capabilities, capable of capturing complex patterns and regularities in time series data. By employing the WaveNet model in the prediction stage, PV output time series data can be better processed, thereby improving prediction accuracy and generalization capability.

Once the model training is completed, the predicted results of different IMF modes are aggregated to obtain the predicted total PV output. The input for this stage is the predicted outputs of different IMF modes, and the output is the aggregated total predicted PV output. This approach combines the different feature IMF time series data and utilizes the model's learning capabilities for each IMF, resulting in a more comprehensive prediction of the total output of the PV system, further improving the accuracy and reliability of the prediction results.

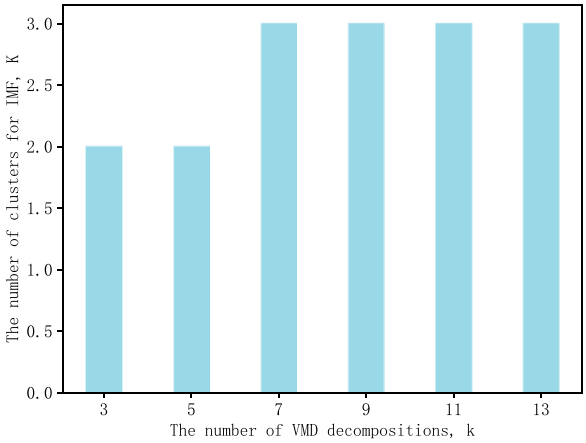


FIGURE 4
Number of clusters for disturbance IMFs clustering at different k.

5 Case study

To validate the effectiveness of the proposed PV output prediction method based on VMD disturbance feature extraction and WaveNet model, historical data from a PV station in Zhejiang, China, was used as the experimental dataset. The historical PV output data covers the fourth quarter of the year 2022, from 1 October 2022, 0:00 to 31 December 2022, 23:55. The data is sampled at a frequency of 5 min per point, and the output data is in units of watts.

5.1 Evaluation metrics

The forecasting part of the experiment focuses on turbine output prediction. The results are evaluated using RMSE, Mean Absolute Error (MAE), and MAPE as the evaluation metrics. The formulas for these metrics are (Equations 9–11):

$$e_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{t=T+1}^{T+N}\left(\alpha_t^s - \hat{\alpha}_t^s\right)^2} \tag{9}$$

$$e_{\mathrm{MAE}} = \frac{1}{N}\sum_{t=T+1}^{T+N}\left|\alpha_t^s - \hat{\alpha}_t^s\right| \tag{10}$$

$$e_{\mathrm{MAPE}} = \frac{1}{N}\sum_{t=T+1}^{T+N}\left|\frac{\alpha_t^s - \hat{\alpha}_t^s}{\alpha_t^s}\right| \tag{11}$$

In power output prediction, $\alpha_t^s = p_t^s$ represents the actual value of the output, while $\hat{\alpha}_t^s = \hat{p}_t^s$ represents the predicted value of the PV output.

5.2 Decomposition of output using VMD

First, the historical output time series is decomposed using VMD to extract and analyze different modes of output variations. In VMD, the parameter α controls the balance between the smoothness of decomposition and the fitting of data. To better fit the data, the

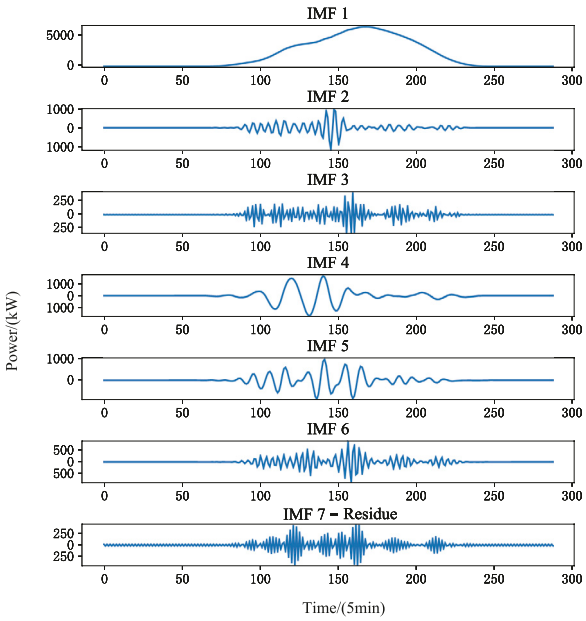


FIGURE 5
The VMD decomposition diagram of PV power output.

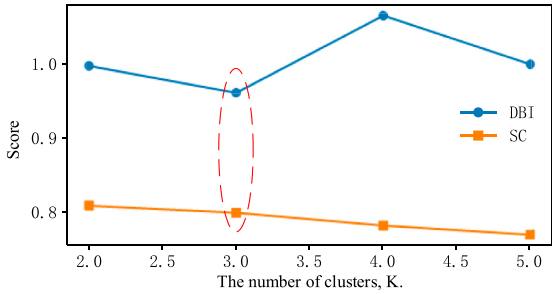


FIGURE 6
Evaluation indices for K-means clustering.

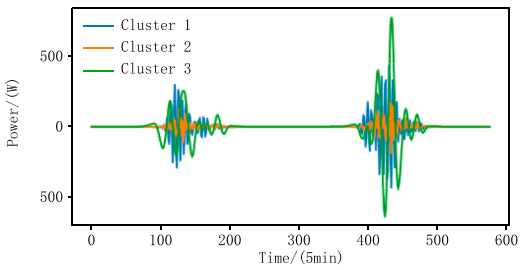


FIGURE 7
The central line of uphill climbing clustering.

experimental setting for the fitting coefficient alpha is set to 10. During the VMD operation, the experiment initially sets the number of modes, k, to 3, 5, 7, 9, 11, etc., and then applies K-means clustering to the obtained disturbance IMFs for different values of k. The number of clustered disturbance IMFs for different values of k is shown in Figure 4.

From Figure 4, it can be observed that as the number of VMD decompositions, k, increases, the subsequent clustering numbers stabilize starting from k = 7. It can be seen that increasing the number of decompositions does not change or improve the clustering effect. Therefore, in this experiment, the number of modes, k, is set to 7. The VMD decomposition diagram of the PV output is shown in Figure 5.

According to Figure 5, VMD decomposes the output into seven characteristic IMFs. IMF1 represents the clear-sky curve of the day, reflecting the output curve unaffected by weather conditions. IMF4 and IMF5 represent high-amplitude low-frequency disturbances caused by changes in cloud cover, while IMF2 and IMF6 represent high-amplitude mid-frequency disturbances caused by changes in cloud cover. IMF3 and IMF7 represent low-amplitude high-frequency disturbances caused by changes in cloud cover and the PV system itself. The Residue represents the noise component of the PV system.

5.3 Disturbance IMF clustering

Next, K-means clustering was applied to the disturbance IMF modes excluding the clear-sky IMF. The clustering SC and DBI scores are shown in Figure 6, and the centroids of the clusters are depicted in Figure 7.

According to Figure 6, when K = 3, both the DBI index and SC index are optimal, indicating the best clustering effect. That is, the disturbance IMF is mainly divided into three categories, and the IMF time series data of the same category are used to construct the prediction model for training in subsequent predictions. Figure 7 shows the cluster centroids of the disturbance IMF, and the clustering results are consistent with the conclusions of VMD decomposition. By identifying different disturbance feature classes, the variation patterns and periodic influences of PV power output disturbances can be explored. Combined with historical meteorological data, further analysis of weather change patterns can improve the accuracy of PV power output prediction.

5.4 Spearman correlation analysis

After obtaining the clustering results from K-means, the Spearman correlation analysis was conducted between the class clear-sky IMF curve and the clustering center lines of disturbance clusters 1, 2, and 3, respectively, with historical weather data. The correlation analysis results are shown in Table 1.

From the data in Table 1, it can be analyzed that there is a very high Spearman correlation between the clear sky IMF and irradiance, indicating a significant correlation between the clear sky IMF and irradiance. However, the Spearman correlations between the disturbance IMF clusters 1, 2, and 3 and various meteorological parameters are relatively low, suggesting weak associations between them and the meteorological parameters. Specifically, disturbance IMF cluster 2 has the lowest Spearman correlation coefficients with all meteorological parameters, indicating the weakest connection with each meteorological parameter. In contrast, disturbance IMF cluster 1 shows relatively

TABLE 1 The results of Spearman analysis.

	Irradiance	Pressure	Surface temperature	Humidity	Wind speed at 70 m
clear-sky IMF	0.9002	0.1907	0.4905	0.7965	0.3008
clusters 1	0.0367	0.0449	0.0063	0.0279	0.0131
clusters 2	0.0153	0.0166	0.0029	0.0242	0.0101
clusters 3	0.4094	0.1705	0.3072	0.2609	0.0398

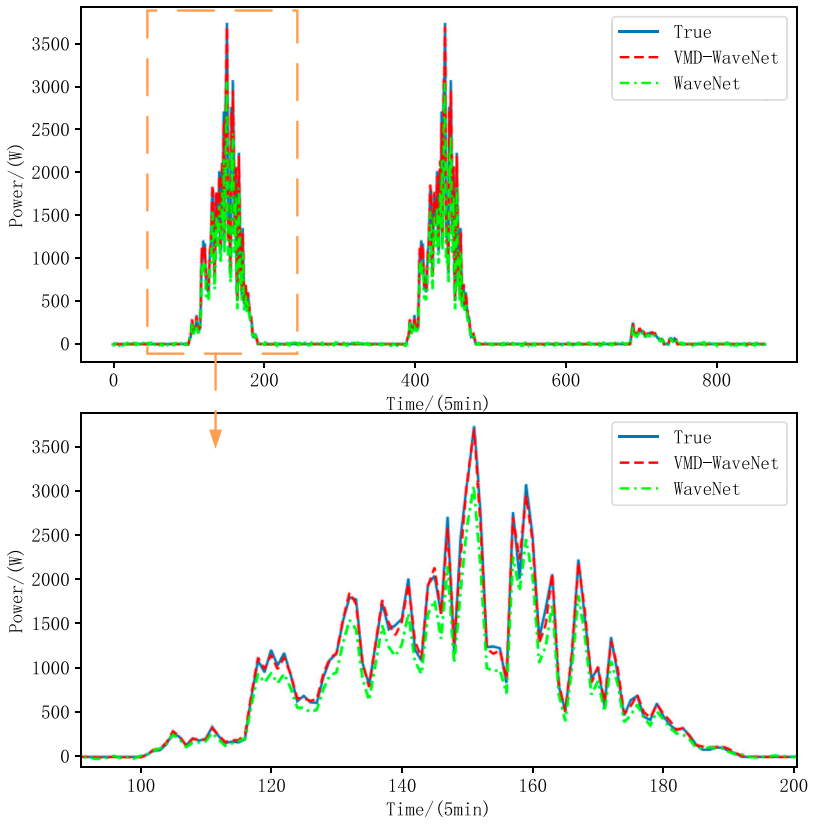


FIGURE 8
Comparison chart of predicted actual values vs. predicted values.

high Spearman correlations with irradiance and surface temperature, while disturbance IMF cluster 3 exhibits relatively high correlation with humidity, indicating that meteorological factors have the greatest influence on disturbance IMF cluster 3. Based on this analysis, this study selects the optimal meteorological factors for each power cluster to construct the experimental dataset.

5.5 WaveNet prediction model

During the prediction phase, WaveNet prediction models were separately constructed for the class-sunny-day IMF and the power IMFs within each disturbance IMF cluster. Predictions were made for each IMF component, and these predictions were then aggregated to obtain the overall PV power output prediction. In the experiment, the above steps were organized into an experimental dataset. The dataset was divided into training and testing sets, with the last 3 days' data reserved for testing and the remaining data used for training.

In the comparative experiment, the models were divided into three categories: CNN-based models, RNN-based models, and hybrid models. The reason for selecting these three categories is to comprehensively evaluate the performance of different types of neural networks in PV power output prediction. CNN-based models excel at handling short-term complex fluctuations and can quickly capture local features in time series data; RNN-based models have advantages in dealing with long-term dependencies and can better capture long-term trends in time series data; hybrid models combine the strengths of both CNNs and RNNs, enabling them to handle both short-term fluctuations and long-term trends. Each category included versions with and without VMD decomposition and Spearman correlation analysis. The specific models included: VMD-TCN, VMD-LSTM, VMD-GRU, VMD-CNN-LSTM,

TABLE 2 Evaluation metrics for prediction results.

Group	Model	RMSE /(W)	MAE /(W)	MAPE /(%)	Model	RMSE /(W)	MAE /(W)	MAPE /(%)
CNN-based	proposed	27.01	10.90	6.94	WaveNet	75.22	32.69	7.17
CNN-based	VMD-TCN	47.41	20.34	6.99	TCN	121.86	45.67	7.32
RNN-based	VMD-LSTM	88.56	35.89	7.28	LSTM	183.04	68.42	7.42
RNN-based	VMD-GRU	76.93	30.56	7.25	GRU	169.77	62.43	7.39
Hybrid models	VMD-CNN-LSTM	70.52	24.76	7.22	CNN-LSTM	152.69	58.27	7.36
Hybrid models	VMD-CNN-GRU	44.93	15.68	7.10	CNN-GRU	141.28	54.32	7.34
Hybrid models	VMD-Transformer	42.09	14.76	7.02	Transformer	132.31	50.89	7.33

VMD-CNN-GRU, VMD-Transformer, and their counterparts without decomposition. Additionally, this study proposed the VMD-WaveNet prediction model. The performance of these models was evaluated by comparing the actual and predicted outputs, as shown in Figure 8, analyzing the performance of different methods in PV power output prediction.

Figure 8 shows the comparison between the actual PV power output (blue solid line) and the predictions from two models: VMD-WaveNet (red dashed line) and WaveNet (green dashed line) over 3 days. The VMD-WaveNet model is closer to the actual values most of the time, especially at the two main peaks around 150 min and 430 min, where it captures the fluctuations in the actual output more accurately, while the WaveNet model shows larger errors at these peaks. The zoomed-in plot further illustrates the details within the 100 to 200 min time period, where the VMD-WaveNet model is closer to the actual values than the WaveNet model, particularly during periods of large fluctuations, demonstrating higher predictive accuracy. At several peaks and troughs, the VMD-WaveNet model better tracks the changes in actual output, whereas the WaveNet model exhibits greater prediction errors at these points. Overall, the VMD-WaveNet model outperforms the WaveNet model in capturing both the overall trend and local fluctuations in PV power output predictions, indicating that the incorporation of VMD decomposition and Spearman correlation analysis significantly enhances the performance of the PV output prediction model.

According to the results in Table 2, CNN-based models such as WaveNet perform well in short-term predictions and can quickly respond to rapid changes in PV power.

Traditional LSTM and GRU models have advantages in handling long-term dependencies. However, models without VMD processing exhibit deficiencies in noise handling and feature extraction, leading to lower prediction accuracy. The RMSE of VMD-LSTM is 88.56W, significantly better than the 183.04W of the undecorated LSTM; similarly, VMD-GRU has an RMSE of 76.93W, compared to the 169.77W of the undecorated GRU, demonstrating the effectiveness of VMD decomposition in these models. Hybrid models such as VMD-CNN-LSTM and VMD-Transformer combine the strengths of CNN and RNN, performing well in handling both short-term fluctuations and long-term trends. The RMSE of VMD-Transformer is 42.09W, better than the 132.31W of the undecorated Transformer, further proving the value of VMD processing.

Compared to other RNN, CNN, and hybrid models, the WaveNet model excels in handling time series data. By utilizing the structure of convolutional neural networks, WaveNet performs exceptionally well in dealing with short-term complex fluctuations and long-term dependencies. Although the prediction accuracy of the WaveNet without VMD processing is slightly inferior to that of VMD-WaveNet, its RMSE is still 75.22W. Notably, the RMSE of WaveNet is the lowest among all models without VMD processing: LSTM has an RMSE of 183.04W, GRU has an RMSE of 169.77W, TCN has an RMSE of 121.86W, and Transformer has an RMSE of 132.31W. This indicates that WaveNet excels in capturing short-term fluctuations and long-term dependencies even without VMD processing, surpassing other traditional RNN and hybrid models, highlighting its advantages in time series data processing.

The VMD-WaveNet model combines the advantages of VMD decomposition with WaveNet's powerful time series processing capabilities. By extracting features of different frequencies through VMD decomposition and conducting Spearman correlation analysis with meteorological data, it can more accurately capture the short-term fluctuations and long-term trends of PV power output. Figure 8 shows that the VMD-WaveNet model significantly outperforms the undecorated WaveNet model in predicting multiple peaks and valleys, especially near the main peaks at 150 min and 430 min, where the VMD-WaveNet model is closer to the actual values. Additionally, the evaluation metrics in Table 2 further confirm this, with the VMD-WaveNet model achieving an RMSE of 27.01W, an MAE of 10.90W, and a MAPE of 6.94%, all significantly better than other models. This demonstrates that the VMD-WaveNet model, through more refined feature extraction and comprehensive consideration of multiple relevant factors, significantly improves the accuracy and stability of PV output prediction, showing the best predictive performance.

6 Conclusion

The existing short-term forecasting techniques for PV power face challenges such as low prediction accuracy and weak correlation

between meteorological factors and power fluctuations. To address these issues, this paper proposes a PV power prediction method based on VMD for disturbance feature extraction and WaveNet model.

To capture the diverse features of PV power output, VMD is applied to decompose the PV power time series into IMFs representing disturbance and clear-sky components. The clear-sky curve reflects regular variations in PV output, while the disturbance curve reflects irregular fluctuations caused by changes in meteorological conditions.

To better understand the impact of different disturbance types on PV output and provide more information and features for the model, the IMFs representing power disturbances are clustered using K-means clustering based on their power change characteristics. Through analysis of these clustered feature clusters and Spearman correlation analysis of weather factors, different types of power fluctuation patterns are explored more accurately, thereby enhancing the predictive performance of the model.

In the prediction stage, a WaveNet model is employed. By combining the corresponding feature IMF time series data with Spearman-correlated meteorological data, a WaveNet model is constructed for training and prediction. The WaveNet model can effectively capture features and patterns in time series data, considering the temporal correlation and non-linear characteristics of the data, thus improving the accuracy and generalization ability of PV power prediction.

In the experimental section, evaluation metrics are computed, and the predicted data from different models are compared with the ground truth data to validate the computational accuracy and effectiveness of the proposed method for PV power prediction. The results demonstrate that the model provides effective predictions of PV power output, thereby supporting operational management of PV stations.

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

SZ: Conceptualization, Data curation, Methodology, Software, Writing–original draft, Writing–review and editing. XY: Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Validation, Writing–original draft. KL: Project administration, Software, Validation, Writing–original draft. XL: Formal Analysis, Resources, Validation, Supervision, Writing–original draft, Writing–review and editing. WQ: Formal Analysis, Funding acquisition, Investigation, Resources, Supervision, Visualization, Writing–original draft, Writing–review and editing. XH: Conceptualization, Validation, Investigation, Software, 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 Zhejiang Electric Power Co., Ltd. (B311SX23000C). 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 SZ, XY, KL, XL, WQ, and XH were employed by State Grid Zhejiang Electric Power Co., Ltd., Shaoxing 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.

References

Bian, H., and Sun, J. (2021). Prediction model of photovoltaic power generation based on typical meteorological weeks and GRNN. *Electr. Power Eng. Technol.* 40 (05), 94–99.

Ceyhun, Y., and Hakan, A. (2021). A kernel extreme learning machine-based neural network to forecast very short-term power output of an on-grid photovoltaic power plant. *Energy Sources, Part A Recovery, Util. Environ. Eff.* 43 (4), 395–412. doi:10.1080/15567036.2020.1801899

Cheng, L., Zang, H., Wei, Z., Ding, T., and Sun, G. (2022). Solar power prediction based on satellite measurements – a graphical learning method for tracking cloud motion. *IEEE Trans. Power Syst.* 37 (3), 2335–2345. doi:10.1109/tpwrs.2021.3119338

Cheng, Y., Zhuang, F., and Xu, W. (2023). Short-term photovoltaic power output prediction based on improved extreme learning machine. *Mod. Electr. Power* 40 (05), 679–686.

Deng, Y., Wang, L., Jia, H., Zhang, X., and Tong, X. (2022). A deep learning method based on bidirectional wavenet for voltage sag state estimation via limited monitors in power system. *Energies* 15 (6), 2273. doi:10.3390/en15062273

Dong, M., Li, X., and Yang, Z. (2024). Research progress on data-driven distributed photovoltaic power prediction methods. *Electr. Power Netw. Clean Energy* 40 (01), 8–28.

Dong, Z., Zheng, L., and Su, R. (2023). A short-term photovoltaic power output prediction method based on IGWO-SNN. *Power Syst. Prot. Control* 51 (01), 131–138.

Jin, W., Lu, L., and Lai, H. (2024). Photovoltaic power prediction based on power characteristics of K-ISSA-LSTM. *Acta Energiae Solaris Sin.* 45 (02), 429–434.

Leiming, S., Tian, P., Shihao, S., Zhang, C., Wang, Y., Fu, Y., et al. (2023). Wind speed prediction by a swarm intelligence based deep learning model via signal decomposition and parameter optimization using improved chimp optimization algorithm. *Energy* 276, 127526. doi:10.1016/j.energy.2023.127526

Meng, A., Xie, Z., Luo, J., Zeng, Y., Xu, X., Li, Y., et al. (2023). An adaptive variational mode decomposition for wind power prediction using convolutional block attention deep learning network. *Energy* 282, 128945. doi:10.1016/j.energy.2023.128945

Miao, L., Li, Q., and Jiang, Y. (2023). Application of deep learning in power system forecasting. *J. Eng. Sci.* 45 (04), 663–672.

Muqaddas, E., Muhammad, H. A., and ChulHwan, K. (2022). An improved partial shading detection strategy based on chimp optimization algorithm to find global maximum power point of solar array system. *Energies* 15 (4), 1549. doi:10.3390/en15041549

Parri, S., Teeparthi, K., and Kosana, V. (2024). A hybrid methodology using VMD and disentangled features for wind speed forecasting. *Energy* 288, 129824. doi:10.1016/j.energy.2023.129824

Pramono, S. H., Rohmatillah, M., Maulana, E., Hasanah, R. N., and Hario, F. (2019). Deep learning-based short-term load forecasting for supporting demand response program in hybrid energy system. *Energies* 12 (17), 3359. doi:10.3390/en12173359

Raiker, G. A., Loganathan, U., and Reddy B., S. (2021). Current control of boost converter for PV interface with momentum-based perturb and observe MPPT. *IEEE Trans. Industry Appl.* 57 (4), 4071–4079. doi:10.1109/tia.2021.3081519

Sleiman, A., and Su, W. (2024). Combined K-means clustering with neural networks methods for PV short-term generation load forecasting in electric utilities. *Energies* 17 (6), 1433. doi:10.3390/en17061433

Wang, H., Peng, C., Liao, B., Cao, X., and Li, S. (2023a). Wind power forecasting based on WaveNet and multitask learning. *Sustainability* 15 (14), 10816. doi:10.3390/su151410816

Wang, X., and Ma, W. (2024). A hybrid deep learning model with an optimal strategy based on improved VMD and transformer for short-term photovoltaic power forecasting. *Energy* 295, 131071. doi:10.1016/j.energy.2024.131071

Wang, Y., Chen, T., Zhou, S., Zhang, F., Zou, R., and Hu, Q. (2023b). An improved Wavenet network for multi-step-ahead wind energy forecasting. *Energy Convers. Manag.* 278, 116709. doi:10.1016/j.enconman.2023.116709

Wang, Y., Ni, A., and Zhu, L. (2022). Research on ultra-short-term photovoltaic power output prediction based on SSA-LERNN. *Control Eng.* 29 (11), 1941–1947.

Wu, H., Shi, M., and Zheng, H. (2023). Short-term interval prediction of photovoltaic power based on EEMD-ALOCO-SVM model. *J. Sol. Energy* 44 (11), 64–71.

Yagang, Z., Zhiya, P., and Hui, W. (2024). Achieving wind power and photovoltaic power prediction: an intelligent prediction system based on a deep learning approach. *Energy* 283.

Zhang, C., Lin, G., and Kuang, Y. (2023a). Short-term photovoltaic power interval prediction based on MEEMD-QUATRE-BILSTM. *Acta Energiae Solaris Sin.* 44 (11), 40–54.

Zhang, J., Li, F., and Wang, T. (2024). A load forecasting method using memory neural network and curve shape correction. *Electr. Power Eng. Technol.* 43 (01), 117–126.

Zhang, M., Zhen, Z., Liu, N., Zhao, H., Sun, Y., Feng, C., et al. (2023b). Optimal graph structure based short-term solar PV power forecasting method considering surrounding spatio-temporal correlations. *IEEE Trans. Industry Appl.* 59 (1), 345–357. doi:10.1109/tia.2022.3213008

Zhang, X., Li, Y., Li, T., Gui, Y., Sun, Q., and Gao, D. W. (2023c). Digital twin empowered PV power prediction. *J. Mod. Power Syst. Clean Energy*. doi:10.35833/MPCE.2023.000351

Zhu, Y., Gu, J., and Meng, L. (2020). Photovoltaic power prediction model based on EMD-LSTM. *Electr. Power Eng. Technol.* 39 (02), 51–58.

Zhuang, S., Gong, X., and Lin, C. (2019). Estimation of daily total radiation exposure based on generalized regression neural network. *Acta Energiae Solaris Sin.* 40 (01), 11–16.

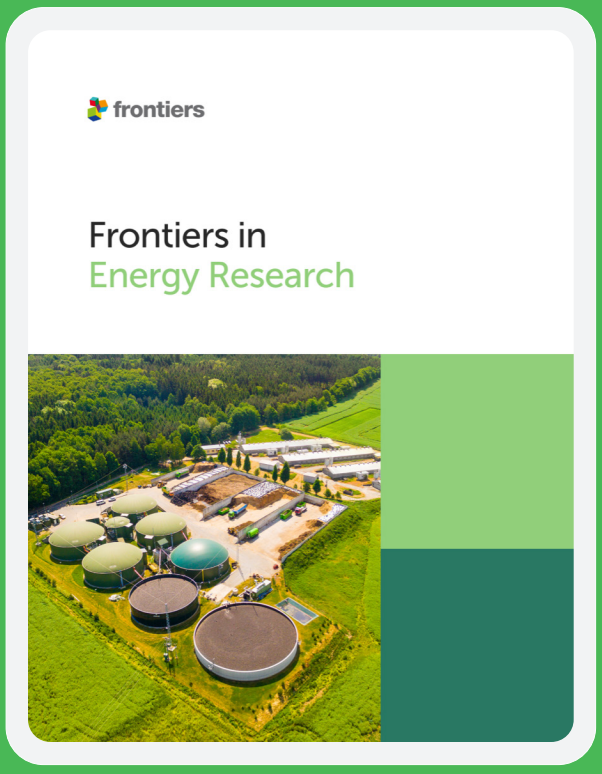
Frontiers in Energy Research

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