

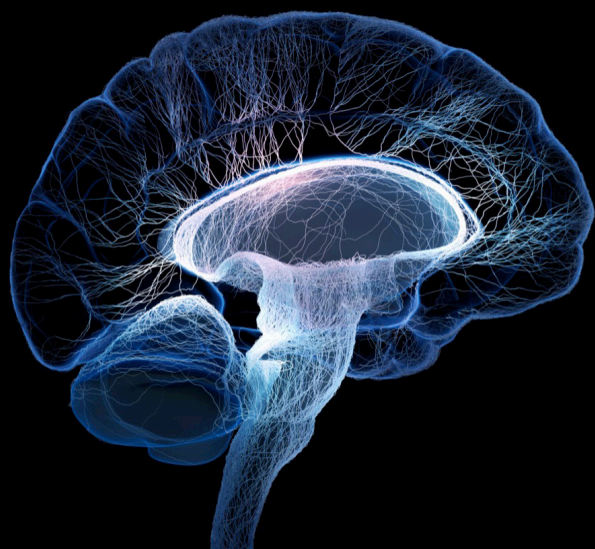
Generative AI for brain imaging and brain network construction

Edited by

Shuqiang Wang, Fei He, Yong Hu and Zhiguo Zhang

Published in

Frontiers in Neuroscience



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 978-2-8325-3507-3
DOI 10.3389/978-2-8325-3507-3

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

Generative AI for brain imaging and brain network construction

Topic editors

Shuqiang Wang — Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (CAS), China

Fei He — Coventry University, United Kingdom

Yong Hu — The University of Hong Kong, Hong Kong, SAR China

Zhiguo Zhang — Harbin Institute of Technology, Shenzhen, China

Citation

Wang, S., He, F., Hu, Y., Zhang, Z., eds. (2023). *Generative AI for brain imaging and brain network construction*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3507-3

Table of contents

- 04 **Editorial: Generative AI for brain imaging and brain network construction**
Shu-Qiang Wang, Zhiguo Zhang, Fei He and Yong Hu
- 07 **Constructing brain functional network by Adversarial Temporal-Spatial Aligned Transformer for early AD analysis**
Qiankun Zuo, Libin Lu, Lin Wang, Jiahui Zuo and Tao Ouyang
- 21 **How far is brain-inspired artificial intelligence away from brain?**
Yucan Chen, Zhengde Wei, Huixing Gou, Haiyi Liu, Li Gao, Xiaosong He and Xiaochu Zhang
- 29 **Prior knowledge-based precise diagnosis of blend sign from head computed tomography**
Chen Wang, Jiefu Yu, Jiang Zhong, Shuai Han, Yafei Qi, Bin Fang and Xue Li
- 40 **Focal cross transformer: multi-view brain tumor segmentation model based on cross window and focal self-attention**
Li Zongren, Wushouer Silamu, Feng Shurui and Yan Guanghui
- 52 **Improved HHT-microstate analysis of EEG in nicotine addicts**
Xin Xiong, Jiannan Feng, Yaru Zhang, Di Wu, Sanli Yi, Chunwu Wang, Ruixiang Liu and Jianfeng He
- 68 **Computational approaches for the reconstruction of optic nerve fibers along the visual pathway from medical images: a comprehensive review**
Richu Jin, Yongning Cai, Shiyang Zhang, Ting Yang, Haibo Feng, Hongyang Jiang, Xiaoqing Zhang, Yan Hu and Jiang Liu
- 81 **Generative AI for brain image computing and brain network computing: a review**
Changwei Gong, Changhong Jing, Xuhang Chen, Chi Man Pun, Guoli Huang, Ashirbani Saha, Martin Nieuwoudt, Han-Xiong Li, Yong Hu and Shuqiang Wang
- 99 **BNLoop-GAN: a multi-loop generative adversarial model on brain network learning to classify Alzheimer's disease**
Yu Cao, Hongzhi Kuai, Peipeng Liang, Jeng-Shyang Pan, Jianzhuo Yan and Ning Zhong
- 110 **Brain organoid data synthesis and evaluation**
Clara Brémond-Martin, Camille Simon-Chane, Cédric Clouchoux and Aymeric Histace



OPEN ACCESS

EDITED AND REVIEWED BY
Vince D. Calhoun,
Georgia State University, United States

*CORRESPONDENCE
Yong Hu
✉ yhud@hku.hk

RECEIVED 18 August 2023
ACCEPTED 28 August 2023
PUBLISHED 05 September 2023

CITATION
Wang S-Q, Zhang Z, He F and Hu Y (2023)
Editorial: Generative AI for brain imaging and
brain network construction.
Front. Neurosci. 17:1279470.
doi: 10.3389/fnins.2023.1279470

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

Editorial: Generative AI for brain imaging and brain network construction

Shu-Qiang Wang¹, Zhiguo Zhang², Fei He³ and Yong Hu^{4*}

¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China,

²Institute of Computing and Intelligence, Harbin Institute of Technology, Shenzhen, China, ³Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry, United Kingdom,

⁴Department of Orthopaedics and Traumatology, The University of Hong Kong, Hong Kong, Hong Kong SAR, China

KEYWORDS

generative artificial intelligence, brain data generation, brain network, brain-inspired computing, diffusion model

Editorial on the Research Topic

Generative AI for brain imaging and brain network construction

1. Introduction

Brain imaging plays an important role in exploring brain which is claimed to be the most complex thing in the universe. In recent years, there has been remarkable progress in the field of brain imaging techniques. Meanwhile, advancements in brain imaging technologies have greatly enhanced our understanding of brain networks. As an important tool for exploring the relationship between brain structure and function, brain network computation represents the most promising direction for artificial intelligence to achieve breakthroughs and advancements in the field of neuroscience. By employing artificial intelligence algorithms to integrate complementary features from multi-modal brain imaging, it is possible to uncover the connectivity characteristics of neural circuits and establish multi-level mapping brain network models based on function, structure, and organization.

Generative artificial intelligence (AI) has witnessed significant expansion, encompassing the utilization of available data to generate fresh content that exhibits comparable underlying patterns to real-world data. The fusion of these two realms, generative AI and neuroimaging, offers a promising path for delving into diverse domains of brain imaging and brain network computation, specifically in the realms of extracting spatio-temporal brain characteristics and reconstructing the topological connectivity of brain networks. Generative artificial intelligence assists researchers in learning and understanding brain functional mechanisms in a broader feature space under limited sample conditions. It aids researchers in designing efficient fusion methods capable of handling and correlating multimodal data and domain knowledge information. By integrating multimodal brain data with prior knowledge from neuroscience, it achieves complementary synergy between the cooperative semantic information and knowledge rules inherent in different levels and factors of the brain.

This Research Topic contains nine articles that can be broadly classified into three categories: (1) four articles primarily focus on the application of generative artificial intelligence methods for enhancing brain data, (2) three articles investigate the relationship between artificial intelligence and brain mechanisms to explore the functioning principles of the brain, (3) the remaining articles primarily address the application and development of artificial intelligence in the diagnosis of brain disorders.

Generative artificial intelligence is one of the rapidly evolving fields. It has gained significant attention due to its inherent advantages, especially in domains where data scarcity is a challenge. The current construction of functional brain connectivity networks from fMRI relies primarily on toolkits, leading to erroneous estimation of connection strength and suboptimal performance in disease analysis. Zuo et al. proposed a novel Adversarial Temporal-Spatial Aligned Transformer (ATAT) model, which realizes a region-guided feature learning network by taking the volume and location of anatomical brain regions into account. And it can adaptively adjust the boundary features of neighboring regions and capture the global functional connectivity patterns of distant brain regions. The validity and superior performance of the proposed model in early Alzheimer's Disease prediction and progression analysis were verified by conducting experiments on the ADNI dataset. Most artificial intelligence models used for neuroimaging classification tasks have limitations in their learning strategies and lack the ability for incremental learning. Cao et al. constructed the BNLoop-GAN model, which combines multi-modal brain networks and a multiple-loop learning algorithm to enhance the accuracy of prediction tasks for Alzheimer's Disease. This advancement tackles the limitations of current AI models employed in neuroimaging classification, which frequently suffer from the absence of incremental learning capabilities during batch training.

According to the study of Gong et al., this review summarizes the integration of advanced brain imaging techniques and generative AI models, which show promise for extracting spatiotemporal brain features and reconstructing topological connectivity of brain networks. It surveys four classic generative models and their applications in brain image computing and analysis. It also discusses challenges and future directions in utilizing these AI techniques for large-scale brain data analysis to understand brain structure-function relationships, aid diagnosis and treatment of brain diseases, and promote neuroscience research. However, issues like data privacy and individual differences pose challenges that need to be addressed through balancing data usage and model interpretability. There is still controversy surrounding the metrics and validation of synthetic images, and visual assessment can be a time-intensive task. Brémond-Martin et al. used similitude metrics and psychovisual evaluation to validate the quality of synthesized images. They compared the results obtained from these two evaluation methods and further tested the images in a segmentation task. The researchers discovered a correlation between certain metrics and psychovisual decisions, suggesting the potential of using specific combinations of blur metrics as possible alternatives to psychovisual evaluations.

Currently, artificial intelligence and medical imaging techniques advancements hold great significance for the progress of brain science and our understanding of the human brain. Jin et al. provided a comprehensive overview of computational methods for reconstructing optic nerve fibers from medical images in a review. It discusses the clinical importance of optic nerve fiber reconstruction for diagnosing neurological diseases and guiding neurosurgery. They describe two main reconstruction strategies, image segmentation and fiber tracking,

with fiber tracking providing more detailed fiber structures. Both conventional and AI-based approaches are reviewed for each strategy, demonstrating the superior performance of AI-based methods. Xiong et al. designed an improved HHT-Microstate analysis method that combines the improved Hilbert Huang Transformation (HHT) decomposition with microstate analysis to investigate the differences in EEG microstate parameters in each frequency band in nicotine addicts. The method is effective in identifying substance addiction disorders and provides new ideas and insights for the brain research of nicotine addiction. Chen et al. reviewed the interplay between neuroscience and artificial intelligence, from AI's early inspiration from the brain to its evolution and remarkable performance with little neuroscience dependence. However, recent collaborations studying neurobiological explainability of artificial intelligence models reveal they may resemble biological computation despite no explicit neuroscientific modeling. The authors proposed a framework to evaluate brain-likeness of artificial intelligence models to enable further improvements under the intertwined development of both fields.

With the advancements in artificial intelligence technology and a deeper understanding of human brain mechanisms, artificial intelligence is playing an increasingly crucial role in the diagnosis of brain disorders. Wang et al. implemented a prior knowledge-based precise diagnosis of blend sign network from head computed tomography scans, which achieves accurate diagnosis of hybrid signs in head computed tomography scans by incorporating a priori knowledge and combining auxiliary tasks and self-knowledge distillation strategies. The method demonstrated superior performance in experiments and has the potential to assist physicians in reducing workload and improving efficiency in clinical practice. Zongren et al. came out with a multi-view brain tumor segmentation model based on cross-window and focal self-attention, which achieves excellent performance with high segmentation accuracy while limiting the computational cost by enlarging the receptive field by parallel cross windows and improve global dependence by using local fine-grained and global coarse-grained interactions, which provides a new solution for the field of brain tumor segmentation.

These articles cover a wide variety of topics including brain medical image enhancement, brain network generation and reconstruction, brain mechanisms exploration, brain-inspired computing, and brain disorders assisted diagnosis. These researches enrich the corresponding research fields with insightful methodologies and techniques, and ultimately offering alternative solutions to effectively enhance the robustness, generalization ability, and interpret ability for related tasks.

We hope that our readers will have a delightful experience when reading these excellent articles.

Author contributions

S-QW: Conceptualization, Methodology, Writing—original draft, Writing—review and editing, Project administration. ZZ: Writing—review and editing. FH: Writing—review and editing. YH: Writing—review and editing.

Funding

This work was supported by the National Natural Science Foundations of China under Grants 62172403, the Distinguished Young Scholars Fund of Guangdong under Grant 2021B1515020019, the Excellent Young Scholars of Shenzhen under Grant RCYX20200714114641211, and Shenzhen Key Basic Research Projects under Grant JCYJ20200109115641762.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

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



OPEN ACCESS

EDITED BY

Fei He,
Coventry University, United Kingdom

REVIEWED BY

Jin Hong,
Zhongshan Institute of Changchun
University of Science and Technology,
China
Hongfei Wang,
The University of Hong Kong, Hong
Kong SAR, China

*CORRESPONDENCE

Lin Wang
lin.wang1@siat.ac.cn

SPECIALTY SECTION

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

RECEIVED 02 November 2022

ACCEPTED 10 November 2022

PUBLISHED 28 November 2022

CITATION

Zuo Q, Lu L, Wang L, Zuo J and
Ouyang T (2022) Constructing brain
functional network by Adversarial
Temporal-Spatial Aligned Transformer
for early AD analysis.
Front. Neurosci. 16:1087176.
doi: 10.3389/fnins.2022.1087176

COPYRIGHT

© 2022 Zuo, Lu, Wang, Zuo and
Ouyang. 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.

Constructing brain functional network by Adversarial Temporal-Spatial Aligned Transformer for early AD analysis

Qiankun Zuo^{1,2}, Libin Lu³, Lin Wang^{2,4*}, Jiahui Zuo⁵ and
Tao Ouyang⁶

¹School of Information Engineering, Hubei University of Economics, Wuhan, China, ²CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, and the SIAT Branch, Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, China, ³School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China, ⁴Guangdong-Hong Kong-Macau Joint Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen, China, ⁵State Key Laboratory of Petroleum Resource and Prospecting, and Unconventional Petroleum Research Institute, China University of Petroleum, Beijing, China, ⁶State Key Laboratory of Geomechanics and Geotechnical Engineering, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan, China

Introduction: The brain functional network can describe the spontaneous activity of nerve cells and reveal the subtle abnormal changes associated with brain disease. It has been widely used for analyzing early Alzheimer's disease (AD) and exploring pathological mechanisms. However, the current methods of constructing functional connectivity networks from functional magnetic resonance imaging (fMRI) heavily depend on the software toolboxes, which may lead to errors in connection strength estimation and bad performance in disease analysis because of many subjective settings.

Methods: To solve this problem, in this paper, a novel Adversarial Temporal-Spatial Aligned Transformer (ATAT) model is proposed to automatically map 4D fMRI into functional connectivity network for early AD analysis. By incorporating the volume and location of anatomical brain regions, the region-guided feature learning network can roughly focus on local features for each brain region. Also, the spatial-temporal aligned transformer network is developed to adaptively adjust boundary features of adjacent regions and capture global functional connectivity patterns of distant regions. Furthermore, a multi-channel temporal discriminator is devised to distinguish the joint distributions of the multi-region time series from the generator and the real sample.

Results: Experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) proved the effectiveness and superior performance of the proposed model in early AD prediction and progression analysis.

Discussion: To verify the reliability of the proposed model, the detected important ROIs are compared with clinical studies and show partial consistency. Furthermore, the most significant altered connectivity reflects the main characteristics associated with AD.

Conclusion: Generally, the proposed ATAT provides a new perspective in constructing functional connectivity networks and is able to evaluate the

disease-related changing characteristics at different stages for neuroscience exploration and clinical disease analysis.

KEYWORDS

functional brain connectivity, temporal-spatial transformer alignment, generative adversarial learning, graph convolutional network, early Alzheimer's disease

1. Introduction

Early Alzheimer's disease (AD) includes the following three successive stages: significant memory concern (SMC), early mild cognitive impairment (EMCI), and late mild cognitive impairment (LMCI). AD is a common long-term neurological disorder in the elderly, which is generally connected with the gradual decline in understanding, judgment, memory, and executive ability until complete loss. AD is known as the leading cause of death among old people worldwide (Zhang et al., 2022), and its great harmfulness brings heavy psychological pressure and economic burden to the families of patients. According to literature (Derby, 2020), the number of people suffering from AD and other dementias in the world currently exceeds 50 million, and the aging population further aggravates the rise of the patient population. However, there is no consensus on the pathological mechanism (Yuzwa et al., 2008; Diplas et al., 2018), and many pharmaceutical companies had tried and failed to develop effective drugs to cure AD. Therefore, early detection and timely intervention for AD are the only possible way in slowing down or preventing disease deterioration (Jack et al., 2013). The development of neuroimaging has made the use of non-invasive AD study become the mainstream of current research because of no side effects on patients (Wang et al., 2018c; Grassi et al., 2019; Yu et al., 2021; Alvi et al., 2022; Lei et al., 2022; You et al., 2022). It is very promising for the scientific community to develop effective methods to detect brain disease and assist clinical treatment from medical imaging data (Wang et al., 2022).

The brain functional network (BFN) derived from functional Magnetic Resonance Imaging (fMRI) describes the functional interactions among spatially distributed brain regions. Brain science indicates that abnormal functional connectivity always appears at the early stage of AD (Berron et al., 2020). The BFN can give a universal understanding of neurological symptoms and unravel the pathogenesis of cognitive diseases. As mentioned in Yu et al. (2020) and Zuo et al. (2021a), the whole brain is divided into several Region-of-Interests (ROIs) according to the anatomical template. The BFN is modeled as a graph, where each node represents the ROI and each edge represents the functional connection strength between paired ROIs. The conventional method is to use a software toolbox to construct functional connectivity (FC) and then extract effective features for disease diagnosis. For

example, Kabbara et al. (2018) investigated the abnormal hub patterns associated with patients' cognitive performance by applying graph-theory analysis on the constructed functional connectivity. This work preserved the topological structure and gained better evaluation performance than the feature extraction algorithm (Wang et al., 2017; Zuo et al., 2021b; Yu et al., 2022) in Euclidean space. Considering the complexity of brain neural activities and noisy data preprocessed from the raw fMRI, it is significant for clinicians to investigate more advanced methods for modeling effective BFNs in early AD analysis.

Brain functional network construction by using time series can be divided into two classes: static-based method, and dynamic-based method. The former utilized the whole brain time series of fMRI to bridge links between ROIs for AD analysis. The direct way of constructing a brain functional network is to compute the person's correlation (PC) between any paired brain regions (Wang et al., 2007). To reduce the possible impact of adjacent ROIs, Fransson and Marrelec (2008) employed partial correlations to handle this problem and achieved good performance in characterizing the changes of the default mode network associated with the disease. But the calculation of an inverse matrix usually comes up with multiple solutions, so researchers adopted certain constraints on the partial correlation estimation for a stable solution. For example, the matrix-regularized network was encoded as modularity prior to optimizing sparse brain network and they (Qiao et al., 2016) discovered potential biomarkers for personalized diagnosis. The latter method benefits the temporal changes of brain functional connectivity for capturing subtle transient neural abnormalities and has recently been a hot spot in neurological disease analysis. The direct approach is to generate a sequence of functional networks and designed a fused learning algorithm to jointly estimate the temporal network for early MCI detection (Wee et al., 2016). Furthermore, the work in Gong et al. (2022) treated the functional time series and functional connectivity as the node features and edges respectively, and developed a graph convolutional network (GCN) based model to generate multiple brain networks for characterizing brain temporal community by setting six-time sliding steps. To address the noisy problem of limited volumes in a sliding window, Zhou et al. (2018) proposed a matrix-regularized learning framework to learn sparse and modular high-order connectivity features for MCI classification. Although many studies have been conducted in BFN construction, they mainly rely on some specific

preprocessing in the software toolboxes to obtain temporal features of each ROI. The drawbacks lie in two fields: one is that the multiple parameter settings may lead to different errors from person to person, and another is that a series of processes can consume much time and fall far away from the goal of clinical application.

Recently, data-driven models are capable of mining effective common characteristics from noisy data. It has been widely applied in various fields of medical image analysis, such as disease severe assessment (Wang et al., 2020c), lesion area segmentation (Hong et al., 2022b), health assessment (Wang et al., 2018b), disease detection (Wang et al., 2018a; Yang et al., 2022), image reconstruction (Hu et al., 2020b). To improve disease analysis performance, many advanced machine learning algorithms are designed to extract discriminative and robust features (Zeng et al., 2017; Lei et al., 2018; Hong et al., 2019; Wang et al., 2020b). Compared with the classification performance of traditional Convolutional Neural Networks (CNN) (Wang S.-Q. et al., 2015), the 3D Convolutional Neural Network (C3D) is good at capturing the local spatial features in a three-dimensional volume and has been successfully applied on the cross-modal image synthesis (Hu et al., 2020a) and disease recognition (Wang et al., 2020a). Moreover, the transformer network (Jiang et al., 2021) can model the global relationship between distant sub-patch regions. The ROI-based features can be learned by C3D and transformer in sequence from 4D fMRI data. Besides, Generative adversarial networks (GANs) are regarded as a special case of variational inference (Mo and Wang, 2009; Wang, 2009) and demonstrates impressive performance in matching generated data distributions. The obvious evidence is the success in generating cross-modal medical images (Hu et al., 2019, 2021) and domain adaptation segmentation (Hong et al., 2022a). It can be used as a regularizer to constrain the representation learning for stable and generalizable disease analysis.

Inspired by the above observations, in this paper, a novel Adversarial Temporal-Spatial Aligned Transformer (ATAT) model is proposed to automatically learn brain functional networks from 4D fMRI for detecting early AD. The constructed brain functional networks are also analyzed to identify important ROIs and abnormal connections. The main contributions of this work are as follows: (1) The region-sequence aligned generator (RAG) is developed to first learn rough ROI-based features by incorporating the brain anatomical information, then finely adjust the boundary features of adjacent ROIs to generate ROI time series and connectivity features. It greatly enhances the ROI time series learning and fully explores the spatial-temporal characteristics and connectivity information among the whole brain. (2) The multi-channel temporal discriminator is designed to constrain the learned ROI time series with the empirical samples. It regularizes the generator optimization and makes the connectivity feature more robust. (3) Experimental classification results prove the

effectiveness of our model, and the discovered important ROIs and abnormal connections may be potential biomarkers for early AD diagnosis or treatment.

The rest of this article is organized as follows. Section 2 describes the novel proposed ATAT model for brain functional network construction. The experimental settings and prediction results with competing methods are presented in Section 3. In the Section 4, the reliability and limitations of this work are discussed. Finally, the Section 5 summarizes the main remarks of this paper.

2. Materials and methods

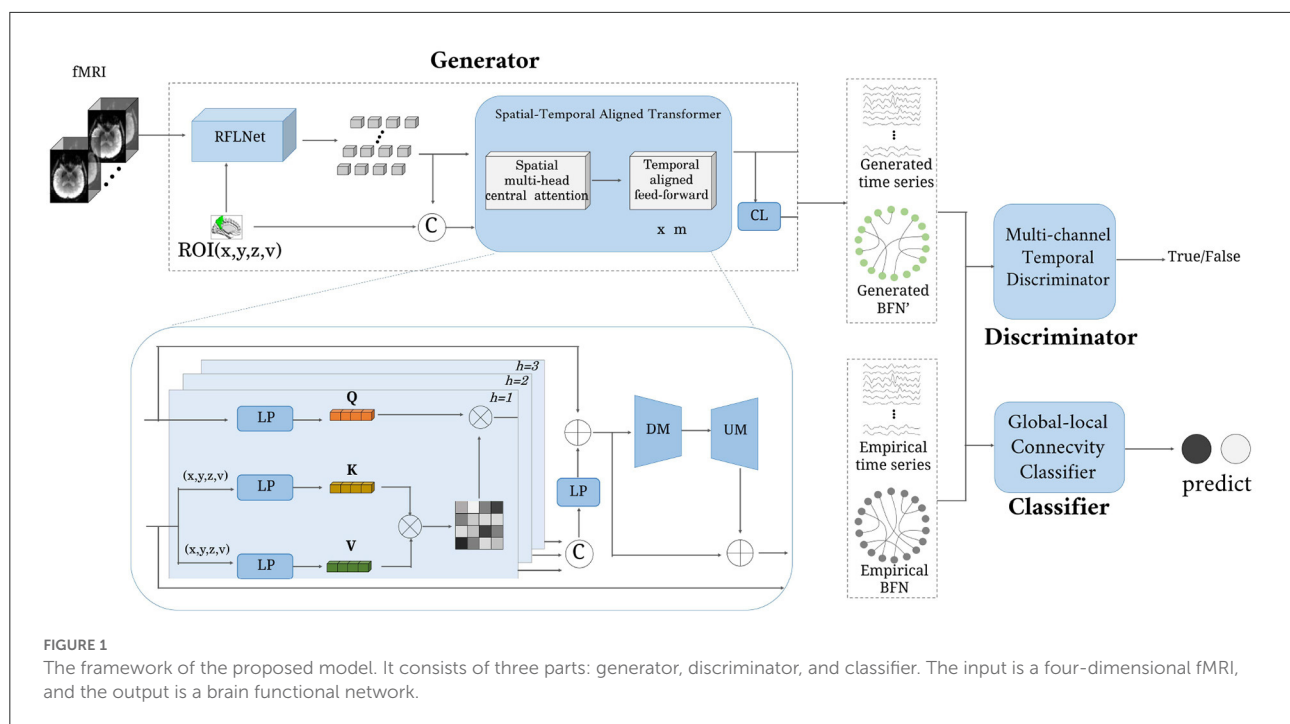
The proposed model includes three main parts, such as (1) data preprocessing, (2) architecture of the proposed model, and (3) objective functions for optimization.

2.1. Data description and preprocessing

The experimental data comes from the public Alzheimer's Disease Neuroimaging Initiative (ADNI-3). A total of 330 subjects with functional Magnetic resonance (fMRI) were downloaded from the website¹, including 86 Normal Control (NC), and three successive stages of early AD (i.e., 82 SMC, 86 EMCI, 76 LMCI). The fMRI data is acquired under the 3.0 Tesla machine. The detailed scanning parameters for fMRI are as follows: the imaging resolution ranges from 2.5 to 3.75 mm along X and Y dimensional direction, the imaging slice thickness is between 2.5 and 3.4 mm; the time of repetition (TR) ranges from 0.607 to 3.0 s, and the time of echo (TE) value is in the range of 30 to 32 ms. The recording time length is about 10 min. The mean age of NC, SMC, EMCI, and LMCI is 74.4, 76.1, 75.7, and 75.8, respectively. The gender is roughly the same in each category.

The fMRI data is preprocessed by the software toolbox GREYNA (Wang J. et al., 2015), which contains about six procedures for constructing ROI-based time series. Each 4D fMRI data is processed by balancing magnetization equilibrium, removing head-motion artifacts, normalizing spatial space, smoothing, and filtering ($0.01\text{Hz} \leq f \leq 0.08\text{Hz}$). Finally, the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) warps the preprocessed image to 90 non-overlapping spatial ROIs, and the final functional features with the size 90×187 are obtained as the truth samples. Meanwhile, the empirical functional connectivity is estimated by calculating the Pearson correlation coefficients between paired ROI time series, and this procedure can generate a 90×90 correlation matrix for each subject.

¹ <http://adni.loni.usc.edu/>



2.2. Architecture

The architecture of the proposed ATAT is shown in Figure 1. It contains three parts: the region-sequence aligned generator (RAG), the multi-channel temporal discriminator (MTD), and the global-local connectivity classifier (GCC). The RAG includes a region-guided feature learning network (RFLNet), and a spatial-temporal aligned transformer (SAT), which transforms the 4D fMRI into ROI time series and brain functional network. Firstly, the raw fMRI data is first sent to the RFLNet for rough ROI-based feature extraction, and the SAT is utilized to finely adjust the feature for adjacent ROIs and align the global temporal correlation between any paired ROIs. Meanwhile, the obtained ROI time series is linearly transformed into brain functional networks through the connectivity learning (CL) network. After that, the generated ROI time series is constrained with the real sample distribution by the MTD. Finally, both ROI time series and brain functional networks are sent to the GCC for disease prediction. There are five objective functions in the model's optimization, including generator loss, discriminator loss, reconstruction loss, classifier loss, and regularized loss.

2.2.1. Region-sequence aligned generator

2.2.1.1. Region-guided feature learning network

As illustrated in Figure 2. This network learns a rough mapping from the raw 4D fMRI to ROI-based time series by introducing the position and volume of the brain anatomical

regions. The size of input data X is $64 \times 64 \times 48 \times 187$. It first passes through four blocks with three successive layers: $3 \times 3 \times 3$ convolutional layers with 1-stride, $2 \times 2 \times 2$ average pooling layer with 2-stride, and a combination layer of *batch normalization* (BN) + *ReLU* activation. The channel number of the above four convolutional layers are 8, 16, 32, 64. Then one $1 \times 1 \times 1$ convolutional layer with 1-stride is used to increase the channels for matching the N ROIs, followed by a *sigmoid* activation function.

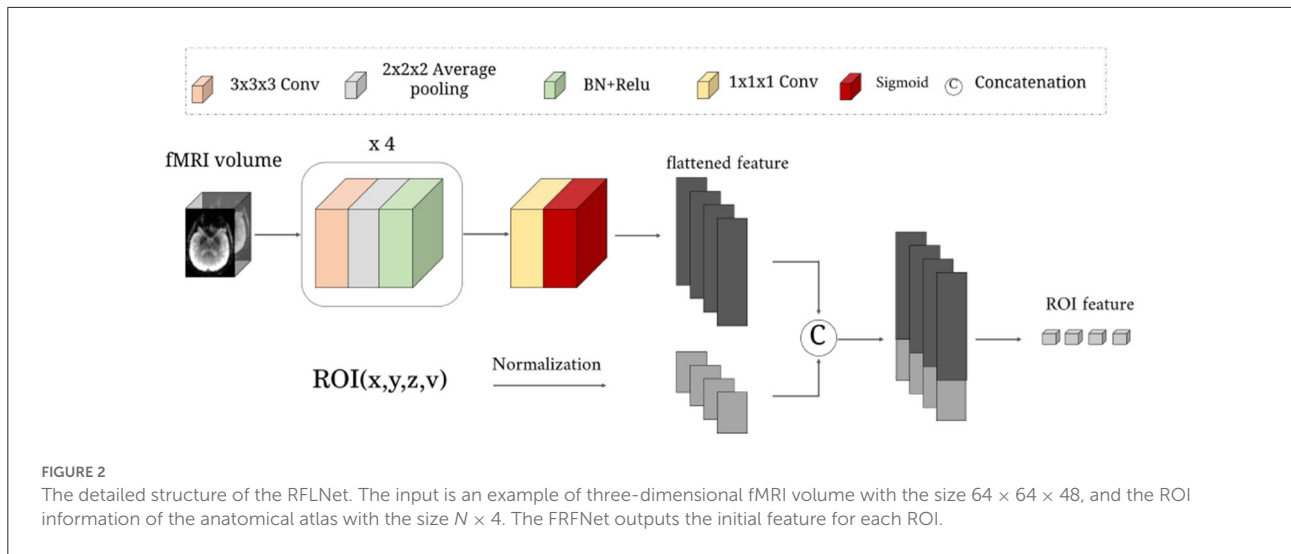
Next, we normalize the central location (x, y, z) and volume (v) of N anatomical ROIs to constrain the brain region information in the range $0 - 1$. Finally, the (x, y, z, v) of N ROIs are treated as ROI embeddings, which are concatenated with the flattened feature of *sigmoid* layer output, which is sent to a one-layer linear projection (LP) layer for generating rough ROI features. The rough ROI feature can be expressed as:

$$F_1 = \text{RFLNet}(X, x, y, z, v) \quad (1)$$

here, the X is the four dimensional volume data fMRI; $x, y, z, v \in \mathbb{R}^{N \times 1}$; $F_1 \in \mathbb{R}^{N \times q}$, the *RFLNet* is a combination of several convolution and pooling operations.

2.2.1.2. Spatial-temporal aligned transformer

To learn more fined ROI temporal features, the spatial-temporal aligned transformer module is designed to recalibrate boundary ROI features and time sequence variations. It splits into two parts: the spatial multi-head central attention (SMCA) and the temporal aligned feed-forward (TAFF). Every ROI is regarded as a token. The rough ROI feature is first sent to three



parallel LP layers to get query (**Q**), key (**K**), and value (**V**). Note that, the calculation of **K** and **V** needs to consider the ROI embeddings. The formulas can be defined as:

$$\mathbf{Q} = LP(F_1), \mathbf{K} = LP(F_1 || x || y || z || v), \mathbf{V} = LP(F_1 || x || y || z || v) \quad (2)$$

where, $||$ means the concatenation operation. Then $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{N \times q}$ are separated into h heads. Each head of token (i.e., $\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i$) has the dimensional size q/h . Taking one head as an example, the central attention (CA) can be expressed:

$$CA_i = \text{Softmax}(\mathbf{Q}_i \mathbf{K}_i^T / \sqrt{q/h}) \mathbf{V}_i \quad (3)$$

here, i means the index of h heads. The output of the spatial multi-head central-attention module is the concatenation of all heads and then with an LP layer (including residual mapping and layer normalization). It can be defined as:

$$SMCA = LP(CA_1 || CA_2 || \dots || CA_h) + F_1 \quad (4)$$

The SMCA has the size $N \times q$.

Next, the TAFF module adjusts the temporal characteristics through the down mapping (DM) and up mapping (UM) layers and reduces the potential noise effect. The DM layer reduces the dimensional of SMCA from q to $q/2$, and the UM layer recovers the feature's dimension. Finally, the output of the TAFF module can be defined as:

$$F_g = UM(DM(SMCA)) + SMCA \quad (5)$$

where, F_g is the generated ROI time series with the size $N \times q$.

To learn an effective brain functional network A_g , we first compute the Euclidean distance between any pair of ROI features and then apply a mapping matrix to it for similarity adjustment. Finally, a Gaussian kernel is introduced to learn

non-linear projection for precise connectivity estimation. The formula can be defined as:

$$A_g(i, j) = \exp\left(-\frac{(F_g^i - F_g^j)^2 W}{2\sigma^2}\right) \quad (6)$$

here, $A_g(i, j)$ represents the functional connectivity between pairwise ROIs. $F_g^i \in \mathbb{R}^{1 \times q}$ means the i th ROI time series. $W \in \mathbb{R}^{q \times q}$ is time series transformable matrix. σ is the bandwidth of the Gaussian kernel, controlling the sparsity with the default value 2.

2.2.2. Multi-channel temporal discriminator

As shown in Figure 3, the multi-channel temporal discriminator (MTD) is used to constrain the generated functional time series (F_g) distribution consistent with the empirical functional time series (F_e). The F_e is computed from the software toolbox, which is treated as the true sample. The structure of MTD consists of N parallel networks, containing three linear projections with $q/2$, q , and $q/2$ neurons. Each MTD accepts i -th ROI time series and outputs one discriminate value. Averaging all the discriminate values is the final discriminate result.

2.2.3. Global-local connectivity classifier

The structure of the global-local connectivity classifier (GCC) is illustrated in Figure 4. It accepts both functional time series (i.e., F_e or F_g) and functional network (i.e., A_e or A_g), outputs the disease label. A total of 5 layers are designed in the GCC, including three graph convolutional layers, one graph pooling layer, and one three-layer perceptron. It is based on the graph convolutional network. The first three layers (i.e., $Gconv_1, Gconv_2, Gconv_3$) are utilized to diffuse

global features and reduce the ROI feature dimension. The graph pooling layer (*Gpool*) is utilized to average features along the ROI feature dimension and get one value for each ROI. And the MLP layer learns a linear mapping to recognize the disease.

2.3. Objective functions

In this section, the five loss functions defined below are utilized to optimize the model for disease prediction and analysis. The reconstruction loss L_{rec} can constrain the generator and retain the empirical features F_e , the generate loss L_g and discriminate loss L_d are combined to optimize the generator and discriminator, the classification loss L_{cls} and regularized loss L_{reg} are utilized to upgrade the parameters of CL network and GCC network. For the convenience of explanation, we make the following simplification: G means all the operations in the Region-sequence aligned generator, D means the multi-channel temporal discriminator, and C is the

global-local connectivity classifier. The raw fMRI data X follows the distribution P_{fMRI} , the P_{F_e} and P_{A_e} represent the empirical functional time series F_e and empirical BFN A_e distribution, respectively. Y is the truth label. These loss functions are defined as follows:

$$L_{rec} = \mathbb{E}_{X \sim P_{fMRI}, F_e \sim P_{F_e}} (||G(X) - F_e||) \quad (7)$$

$$L_g = \mathbb{E}_{X \sim P_{fMRI}} [(1 - D(G(X)))^2] \quad (8)$$

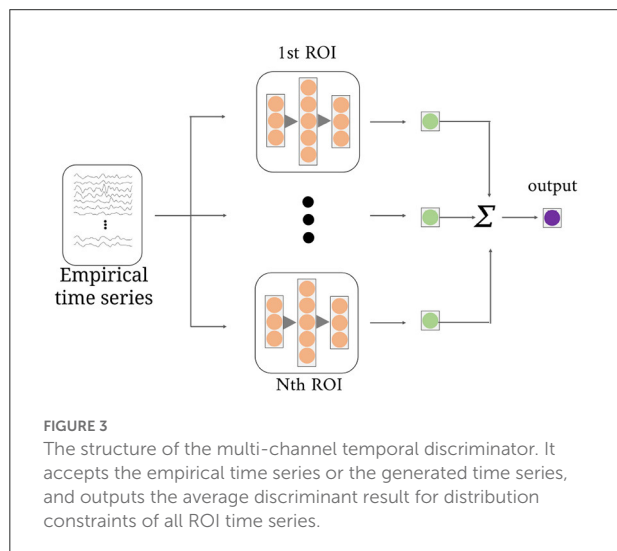
$$L_d = \mathbb{E}_{X \sim P_{fMRI}} [(D(G(X)))^2] + \mathbb{E}_{F_e \sim P_{F_e}} [(1 - D(F_e))^2] \quad (9)$$

$$L_{cls} = \mathbb{E}_{X \sim P_{fMRI}} [-Y \cdot \log(C(G(X)))] + \mathbb{E}_{A_e \sim P_{A_e}, F_e \sim P_{F_e}} [-Y \cdot \log(C(A_e, F_e))] \quad (10)$$

$$L_{reg} = \mathbb{E} (||W||) \quad (11)$$

The hybrid cost of the proposed model is:

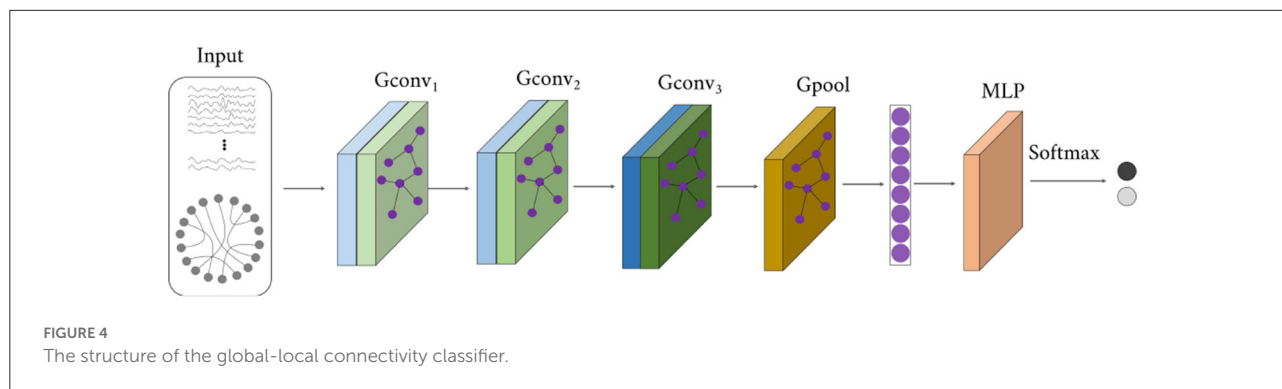
$$L_{all} = L_{rec} + L_g + L_d + L_{cls} + \lambda L_{reg} \quad (12)$$



3. Experiments and results

3.1. Experimental setup

There are six binary classification tasks for the evaluation of the proposed model, including NC vs. SMC, NC vs. EMCI, NC vs. LMCI, SMC vs. EMCI, SMC vs. LMCI, and EMCI vs. LMCI. The evaluation metrics are Accuracy (ACC), Sensitive (SEN), Specificity (SPE), and F1-score. We repeated the 10 times experiment using the five-fold cross-validation on each binary classification and utilized the mean value metrics for the final prediction. To demonstrate our model's good ability in FBN construction, we introduce two classifiers [i.e., SVM (Suthaharan, 2016) and GCN (Kipf and Welling, 2016)] to compare the BFN constructed by ATAT and Empirical.



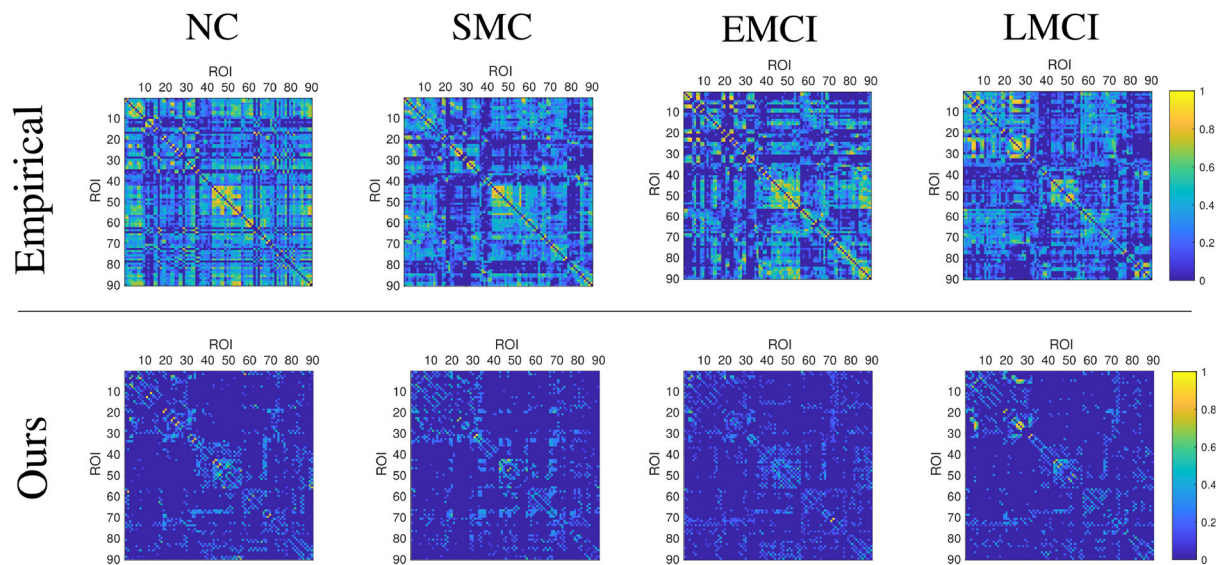


FIGURE 5
Display of brain functional network examples at different disease stages. The BFNs in the upper row are generated by the GREYNA toolbox, and the BFNs in the lower row are generated by the proposed model.

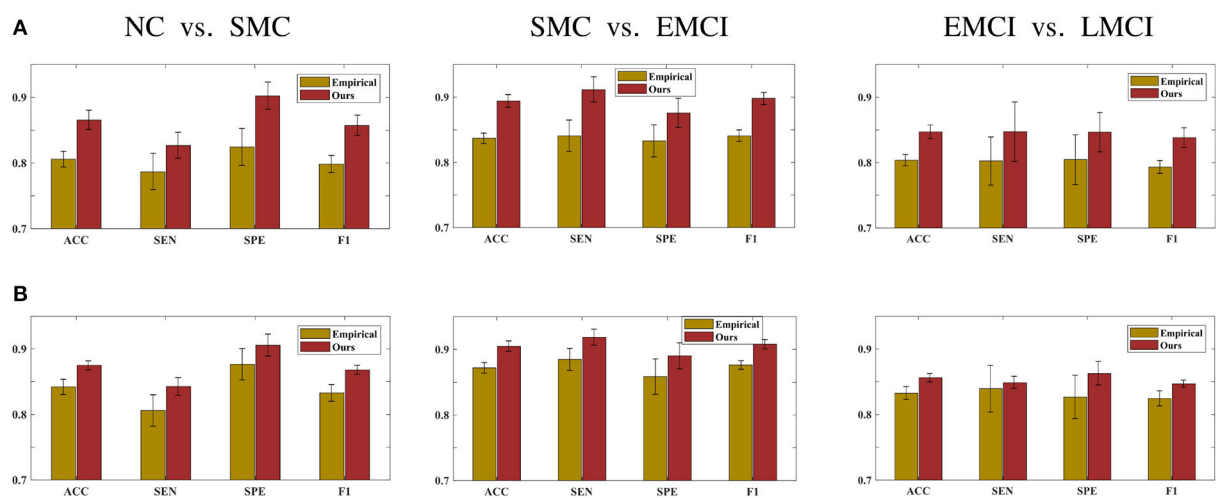


FIGURE 6
Prediction results of three scenarios tasks using (A) the SVM classifier and (B) the GCN classifier.

Our proposed model was implemented with the TensorFlow framework on Ubuntu18.04 and the GPU of NVIDIA GeForce RTX 3080 Ti. The parameters in the experiments are defined as follows: $N = 90, q = 187, h = 11, m = 3, \lambda = 10^{-5}$. During the training, we first update the weights in the generator and the discriminator, then fix part of the generator and optimize the network of CL and GCC. The learning rate of the generator and the classifier were set to 0.0001, and for the discriminator, the learning rate is set to 0.0004. The Adam was adopted for training the proposed model with batch size 2.

3.2. Prediction results

This section demonstrates the good performance of BFN constructed by the proposed model. As shown in Figure 5, the upper row shows the four stages of empirical FBN derived from the GREYNA, while the lower row displays the corresponding BFNs by the proposed model. The main connectivity patterns have been preserved and dense connections become sparse by comparing the empirical and ours. Figure 6 gives the classification result comparison in terms of three scenarios tasks. For the GCN classifier, the BFNs constructed by ours achieve the

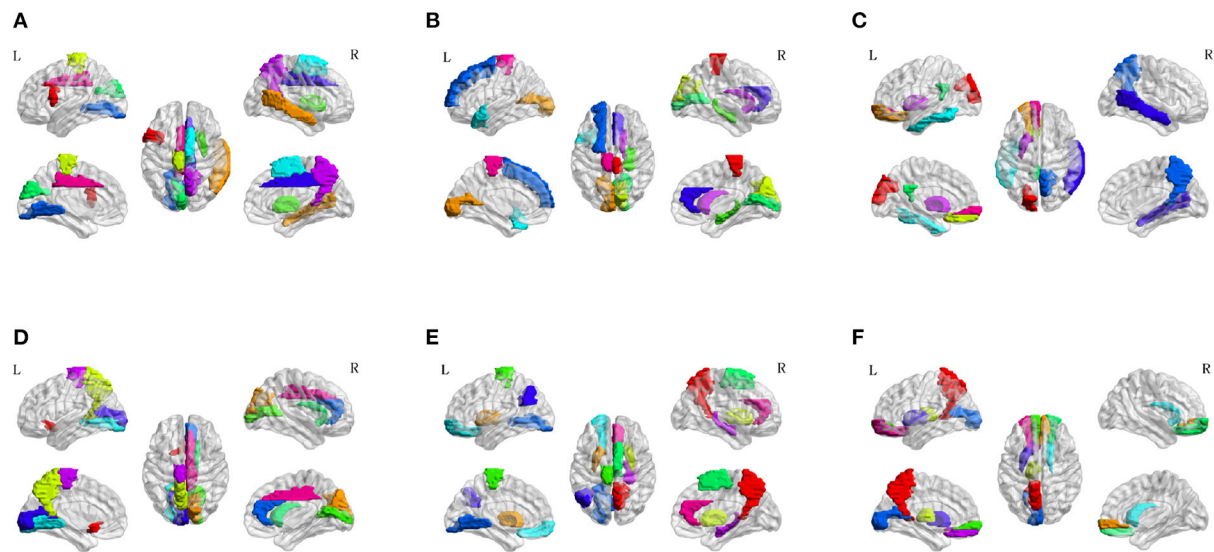


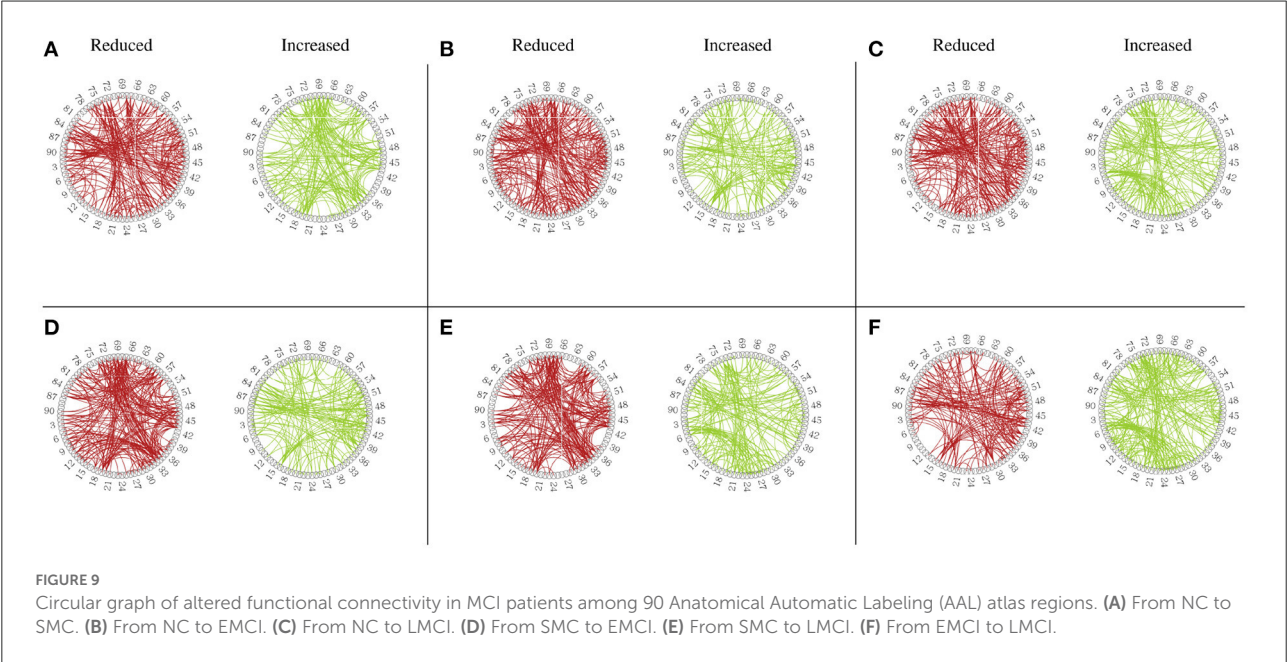
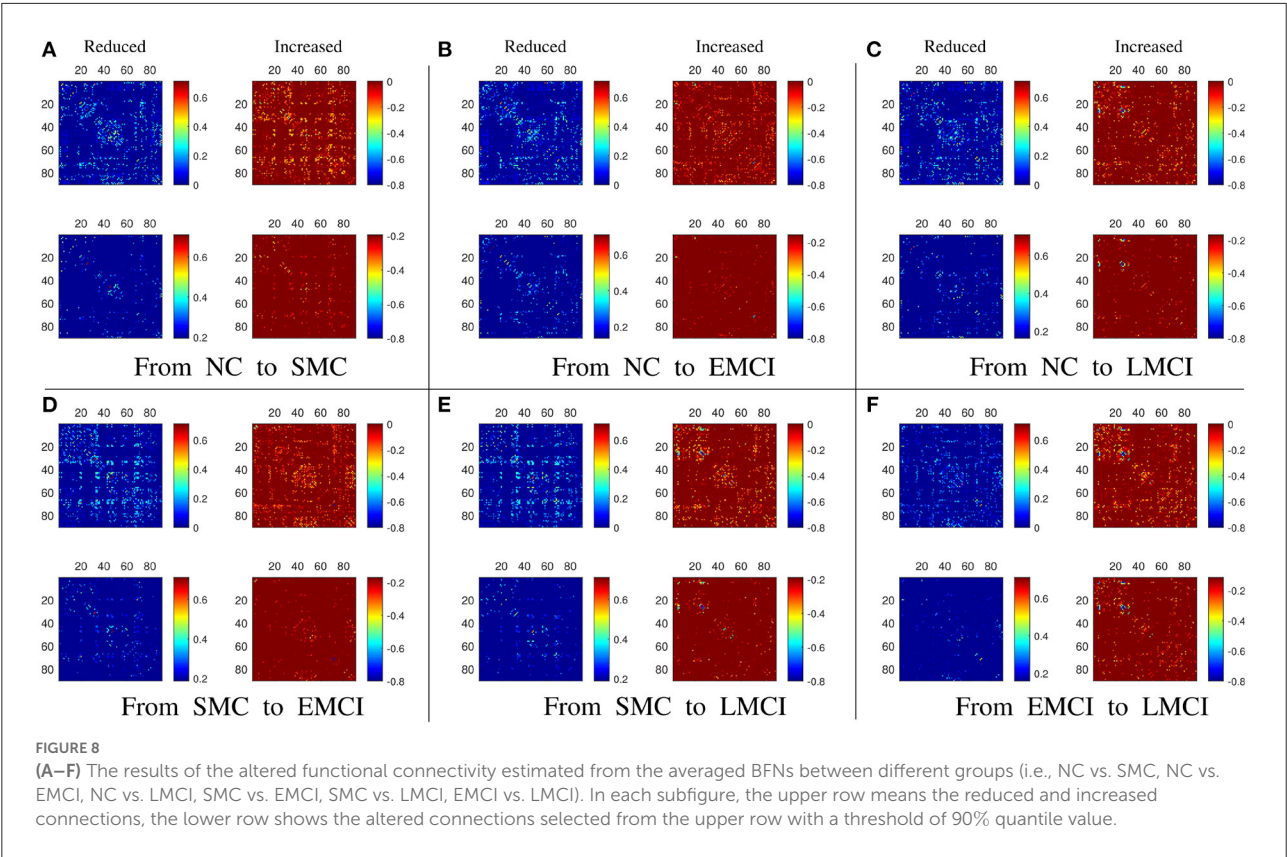
FIGURE 7
Spatial visualization of top 10 brain regions in the six classification scenarios. (A) NC vs. SMC. (B) NC vs. EMCI. (C) NC vs. LMCI. (D) SMC vs. EMCI. (E) SMC vs. LMCI. (F) EMCI vs. LMCI.

best prediction results with a mean ACC of 87.50%, a mean SEN of 84.26%, a mean SPE of 90.58% and a mean F1 of 86.81% in NC vs. SMC task; the mean values of SMC vs. EMCI are 90.47, 91.86, 89.02, and 90.80%; in EMCI vs. LMCI task, the predicted results are 85.61, 84.86, 86.27, and 84.70%. The standard error also shows the superior stability of the proposed model.

To investigate the potential AD-related ROIs, we shield one brain region and calculate the classification ACC as the effect of this ROI on AD progression. After sorting the ACCs in ascending order, the top 10 values are the most important ROI in the classification evaluation. As is shown in Figure 7, the spatial distribution of 10 important AD-related ROIs is displayed in lateral, medial, and dorsal views using the BrainNet Viewer (Xia et al., 2013). Specifically, the top 10 related ROIs are IFGoperc.L, MTG.R, PCL.L, PUT.R, CUN.L, SMA.R, LING.L, DCG.R, PCUN.R, DCG.L in NC vs. SMC classification scenario; The ten ROIs, including PCL.R, CAL.L, CUN.R, HIP.R, CAL.R, TPOsup.L, SFGdor.L, ACG.R, CAU.R, PCL.L, are important for NC vs. EMCI; also, the top 10 ROIs of NC vs. LMCI are SOG.L, ORBsup.L, REC.L, PUT.L, PCG.L, ITG.L, PCUN.R, MTG.R, PUT.L, ORBsupmed.L; For SMC vs. EMCI and SMC vs. LMCI classification, the important ROIs are OLF.L, CUN.R, PCUN.L, CAL.R, CAU.R, LING.L, ACG.R, CAL.L, PCL.L, DCG.R, and PCUN.R, PUT.L, PUT.R, PCL.L, SMA.R, ORBsup.L, LING.L, ANG.L, HIP.R, ACG.R, respectively; For EMCI vs. LMCI, the important ROIs are PCUN.L, ORBsupmed.R, THA.L, ORBsupmed.L, ORBsup.R, CAU.R, CAL.L, PUT.L, REC.L, ORBsup.L.

3.3. Brain network analysis

Besides the prediction of different early AD stages, the other major purpose is to analyze the learned FBNs. After applying the ATAT model to each subject, we can obtain the mean FBN for each group of patients (i.e., NC, SMC, EMCI, and LMCI). To investigate the altered connectivity of FBN between different groups, we compute the difference of six paired scenarios as shown in Figure 8. In each subplot, reduced and increased connectivity can be observed between different paired groups. To analyze the significant connections, we set the 90% quantile value of the altered connectivity strength as the threshold. The pictures in the lower row of each subplot are the corresponding connectivity matrices by setting the threshold value. Figure 9 shows these significant connections in a circular graph. The number of reduced connections are 219, 263, 235, 251, 222, and 163 for NC vs. SMC, NC vs. EMCI, NC vs. LMCI, SMC vs. EMCI, SMC vs. LMCI, EMCI vs. LMCI, respectively; the corresponding number of increased connections are 183, 139, 166, 150, 179, 239. To show the main connectivity patterns in different classification scenarios, we select the top 2% largest altered connections (i.e., reduced, and increased). As shown in Figure 10, different connectivity patterns can be seen in different classification scenarios. Figure 11 depicts the top 5 reduced and top 5 increased connections in the axial and coronal view direction. The connectivity-related ROIs are listed in Tables 1, 2.



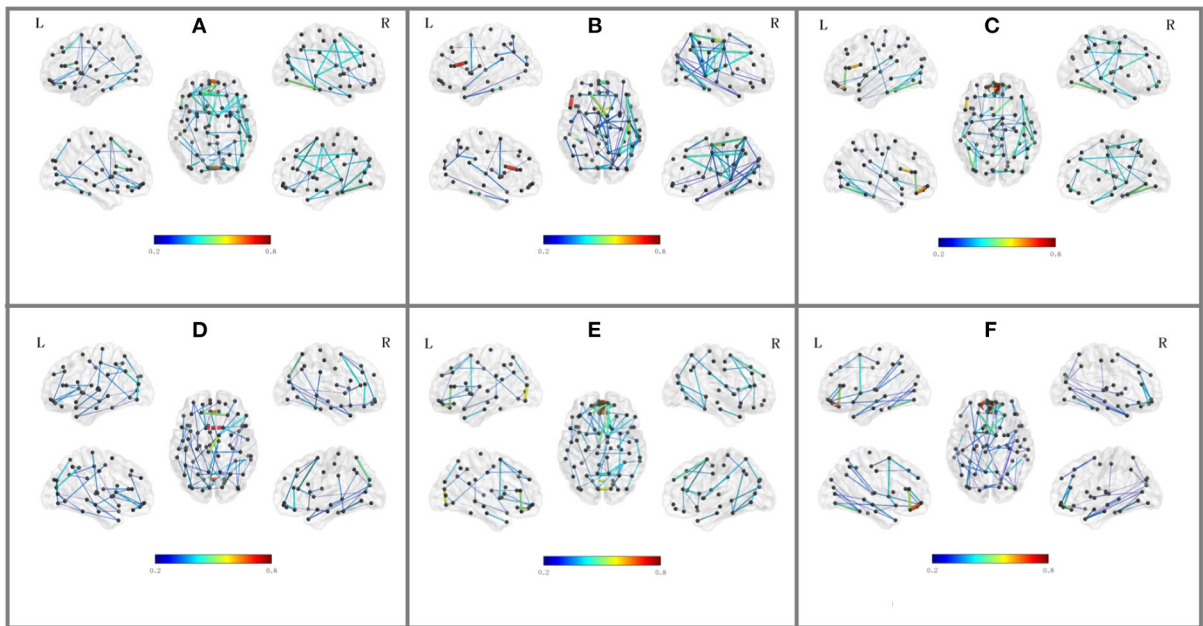


FIGURE 10
Top 2% altered functional connections in strength evaluation in the six classification scenarios. Each subfigure shares the same color bar, which means the absolute connection strength. **(A)** NC vs. SMC. **(B)** NC vs. EMCI. **(C)** NC vs. LMCI. **(D)** SMC vs. EMCI. **(E)** SMC vs. LMCI. **(F)** EMCI vs. LMCI.

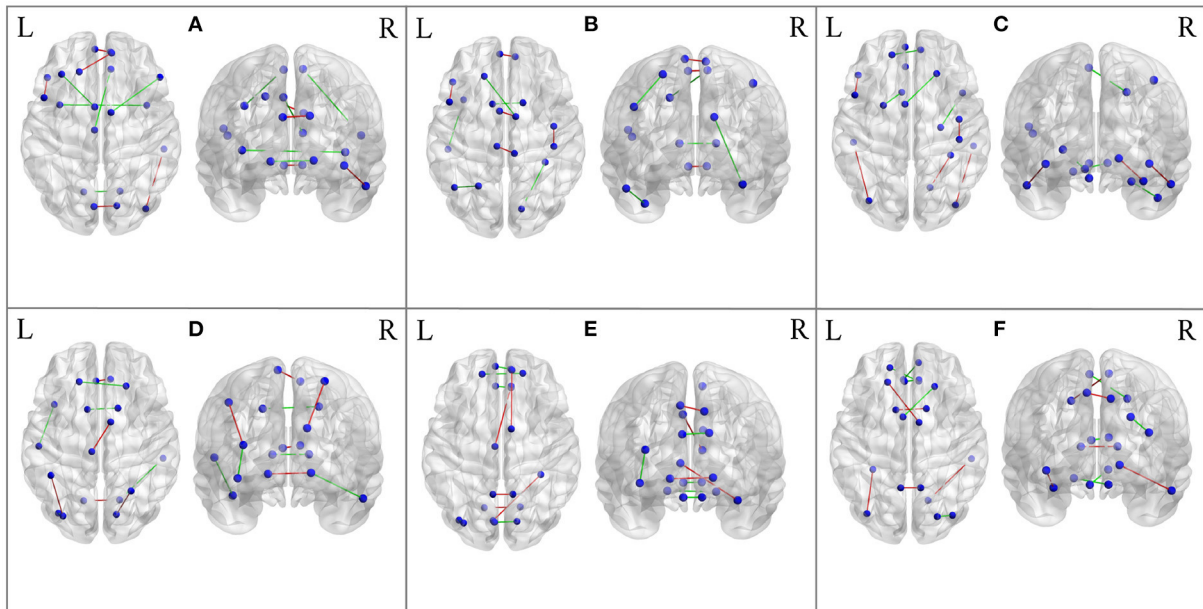


FIGURE 11
The most significant 5 reduced connections and 5 increased connections mapped on the AAL 90 template using the BrainNet Viewer software package. Blue color means the ROIs, red color means reduced connections, and green color means increased connections. **(A)** From NC to SMC. **(B)** From NC to EMCI. **(C)** From NC to LMCI. **(D)** From SMC to EMCI. **(E)** From SMC to LMCI. **(F)** From EMCI to LMCI.

TABLE 1 The top 10 significant altered connections estimated from the generated FBNs in NC vs. SMC, NC vs. EMCI, NC vs. LMCI using AAL90 template (– means reduced connections, + means increased connections).

From NC to SMC		From NC to EMCI		From NC to LMCI	
Indices	Names	Indices	Names	Indices	Names
11, 13	IFGoperc.L, IFGtriang.L	11, 13	IFGoperc.L, IFGtriang.L	11, 13	IFGoperc.L, IFGtriang.L
3, 24	SFGdor.L, SFGmed.R	19, 20	SMA.L, SMA.R	48, 56	LING.R, FFG.R
– 25, 26	ORBsupmed.L, ORBsupmed.R	25, 26	ORBsupmed.L, ORBsupmed.R	2, 58	PreCG.R, PoCG.R
45, 46	CUN.L, CUN.R	2, 58	PreCG.R, PoCG.R	53, 89	IOG.L, ITG.L
54, 90	IOG.R, ITG.R	6 9, 70	PCL.L, PCL.R	54, 90	IOG.R, ITG.R
7, 19	MFG.L, SMA.L	3, 20	SFGdor.L, SMA.R	4, 19	SFGdor.R, SMA.L
14, 20	IFGtriang.R, SMA.R	46, 56	CUN.R, FFG.R	5, 26	ORBsup.L, ORBsupmed.R
+ 29, 30	INS.L, INS.R	59, 65	SPG.L, ANG.L	25, 27	ORBsupmed.L, REC.L
32, 33	ACG.R, DCG.L	71, 72	CAU.L, CAU.R	21, 73	OLF.L, PUT.L
47, 48	LING.L, LING.R	87, 89	TPOmid.L, ITG.L	40, 88	PHG.R, TPOmid.R

TABLE 2 The top 10 significant altered connections estimated from the generated FBNs in SMC vs. EMCI, SMC vs. LMCI, EMCI vs. LMCI using AAL90 template (– means reduced connections, + means increased connections).

From SMC to EMCI		From SMC to LMCI		From EMCI to LMCI	
Indices	Names	Indices	Names	Indices	Names
31, 32	ACG.L, ACG.R	20, 24	SMA.R, SFGmed.R	3, 20	SFGdor.L, SMA.R
47, 48	LING.L, LING.R	32, 33	ACG.R, DCG.L	53, 55	IOG.L, FFG.L
– 46, 60	CUN.R, SPG.R	47, 48	LING.L, LING.R	67, 68	PCUN.L, PCUN.R
51, 61	MOG.L, IPL.L	43, 56	CAL.L, FFG.R	71, 72	CAU.L, CAU.R
20, 69	SMA.R, PCL.L	67, 68	PCUN.L, PCUN.R	48, 90	LING.R, ITG.R
3, 4	SFGdor.L, SFGdor.R	5, 6	ORBsup.L, ORBsup.R	4, 19	SFGdor.R, SMA.L
51, 53	MOG.L, IOG.L	25, 26	ORBsupmed.L, ORBsupmed.R	26, 27	ORBsupmed.R, REC.L
+ 71, 72	CAU.L, CAU.R	27, 28	REC.L, REC.R	5, 28	ORBsup.L, REC.R
81, 83	STG.L, TPOsup.L	45, 46	CUN.L, CUN.R	31, 32	ACG.L, ACG.R
48, 90	LING.R, ITG.R	51, 53	MOG.L, IOG.L	50, 52	SOG.R, MOG.R

4. Discussion

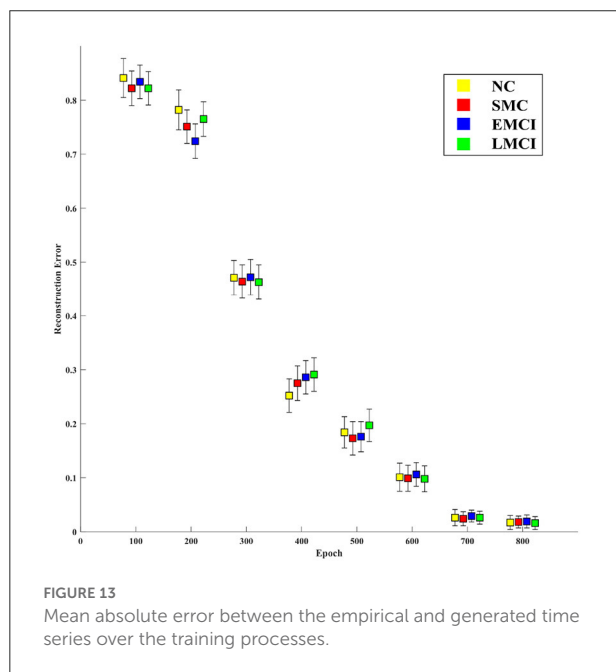
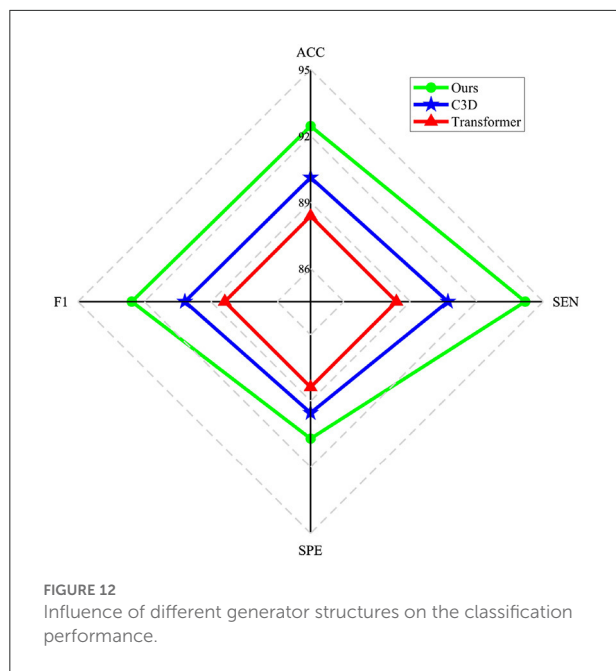
4.1. Effect of the generator

The main goal of the proposed model is to generate BFNs from 4D fMRI data. The modules in the generator play an important role in disease prediction and brain network analysis. To investigate the influence of the generator structure on the classification performance (i.e., NC vs. LMCI), we replace the RFLNet and the SAT modules with traditional C3D (Hong et al., 2020) and transformer (Jiang et al., 2021), respectively. In both cases, the anatomical ROI information is not included in the module. Figure 12 shows that either the C3D or Transformer network can degrade the prediction performance, and the traditional transformer network has a worse influence on classification than the C3D network. It may indicate the proposed RFLNet learns rough ROI-based features with a litter effect on the results, and the SAT network finely adjusts the adjacent ROI-based temporal features which may

greatly influence the classification performance. Furthermore, the reconstructed error of the ROI time series is measured by the mean absolute error (MAE) metric. As shown in Figure 13, the divergence of the MAE for each disease (i.e., NC, SMC, EMCI, and LMCI) demonstrates the reliable results of the designed generator.

4.2. Comparison with related works

In the six classification scenarios, there are eight ROIs that overlap more than three times in the identified brain regions. These important ROIs are the orbital part of the superior frontal gyrus (ORBsup.L), the anterior cingulate and paracingulate gyri (ACG.R), the calcarine fissure and surrounding cortex (CAL.L), the lingual gyrus (LING.L), the precuneus (PCUN.R), the paracentral lobule (PCL.L), the caudate nucleus (CAU.R), the lenticular nucleus putamen (PUT.L). Most of these ROIs are consistent with the previous studies



(Li et al., 2017; Yu et al., 2017; Ye et al., 2019), which demonstrates the strongly correlation with AD. In addition, the most significant altered connectivity (also called abnormal connections) related ROIs contain the identified eight brain regions. These listed ROIs in the table are mainly distributed in the frontal lobe, temporal lobe, and occipital lobe. The frontal lobe is located in the most anterior part of the cerebral hemispheres, accounting for the first 1/3 of the surface of the cerebral hemispheres. It is mainly related to higher mental

activity, including physical activity control, language speaking, self-awareness, and emotional expression. The identified ROIs (i.e., SFGdor, ORBsup, IFGoperc, IFGtriang) associated with the superior frontal gyrus can be founded in Whitwell et al. (2007). The visual and language information is memorized by the temporal lobe, in which patients with AD showed abnormal levels of tau protein in the inferior temporal gyrus (Mormino et al., 2016). The occipital lobe participates in visual processing, for example, the lingual gyrus shows altered functional connectivity in AD patients (Skouras et al., 2019). In general, the derived important ROIs and abnormal connections by the proposed model can reflect the main characteristics associated with AD.

4.3. Limitations and future directions

Although the proposed model in the experiment has achieved good classification results and reliable connectivity analysis, there are two deficiencies in this work. One is that it only considers the binary classification tasks, which may not capture the continuously changing characteristics during the disease progression. We will conduct multi-class classification experiments to investigate the common changes at different stages of AD. Another limitation is that the dataset used in this study is relatively small. In the future study, we will increase the amount of data to validate the proposed model for brain disorder analysis by using other larger datasets [UK biobank (Sudlow et al., 2015), ABIDE (Heinsfeld et al., 2018)].

5. Conclusion

In this paper, we proposed a novel ATAT model to construct brain functional networks for early AD diagnosis and analysis. The three-player generative adversarial network is alternatively optimized and can learn effective functional connectivity features from 4D fMRI. By incorporating the brain anatomical information, the rough ROI features can be extracted by focusing on the local spatial information of individual brain region. Furthermore, the SAT module explores the temporal characteristics and connectivity information for finely adjusting the boundary features of adjacent ROIs. Meanwhile, the generated features from the region-sequence aligned generator are constrained by the adversarial loss and reconstruction loss. Compared to the empirical method, the brain functional networks constructed by the proposed model achieve higher classification performance. The identified important ROIs and abnormal connections may be the potential biomarkers for early AD diagnosis. Generally, our proposed model has the potential in constructing complex functional connectivity features and exploring abnormal functional connections for neurodegenerative diseases study.

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/s.

Author contributions

QZ designed the methodology, conducted the experiment, and wrote the manuscript. LW was the head of the funds of the paper and supervised and verified the reliability of the experimental results. LL provided suggestions for result visualization. JZ and TO reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported in part by National Key R&D Program of China under grant 2020YFC2005803,

References

- Alvi, A. M., Siuly, S., Wang, H., Wang, K., and Whittaker, F. (2022). A deep learning based framework for diagnosis of mild cognitive impairment. *Knowledge Based Syst.* 248, 108815. doi: 10.1016/j.knsys.2022.108815
- Berron, D., van Westen, D., Ossenkoppele, R., Strandberg, O., and Hansson, O. (2020). Medial temporal lobe connectivity and its associations with cognition in early Alzheimer's disease. *Brain* 143, 1233–1248. doi: 10.1093/brain/awaa068
- Derby, C. A. (2020). Trends in the public health significance, definitions of disease, and implications for prevention of Alzheimer's disease. *Curr. Epidemiol. Rep.* 7, 68–76. doi: 10.1007/s40471-020-00231-8
- Diplas, B. H., He, X., Brosnan-Cashman, J. A., Liu, H., Chen, L. H., Wang, Z., et al. (2018). The genomic landscape of tert promoter wildtype-IDH wildtype glioblastoma. *Nat. Commun.* 9, 1–11. doi: 10.1038/s41467-018-04448-6
- Fransson, P., and Marrelec, G. (2008). The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: evidence from a partial correlation network analysis. *Neuroimage* 42, 1178–1184. doi: 10.1016/j.neuroimage.2008.05.059
- Gong, C., Xue, B., Jing, C., He, C.-H., Wu, G.-C., Lei, B., et al. (2022). Time-sequential graph adversarial learning for brain modularity community detection. *Math. Biosci. Eng.* 19, 13276–13293. doi: 10.3934/mbe.2022621
- Grassi, M., Rouleaux, N., Caldirola, D., Loewenstein, D., Schruers, K., Perna, G., et al. (2019). A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. *Front. Neurol.* 10, 756. doi: 10.3389/fneur.2019.00756
- Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., and Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the abide dataset. *Neuroimage Clin.* 17, 16–23. doi: 10.1016/j.nicl.2017.08.017
- Hong, J., Cheng, H., Wang, S.-H., and Liu, J. (2019). Improvement of cerebral microbleeds detection based on discriminative feature learning. *Fund. Inform.* 168, 231–248. doi: 10.3233/FI-2019-1830
- Hong, J., Feng, Z., Wang, S.-H., Peet, A., Zhang, Y.-D., Sun, Y., et al. (2020). Brain age prediction of children using routine brain MR images via deep learning. *Front. Neurol.* 11, 584682. doi: 10.3389/fneur.2020.584682
- Hong, J., Yu, S. C.-H., and Chen, W. (2022a). Unsupervised domain adaptation for cross-modality liver segmentation via joint adversarial learning and self-learning. *Appl. Soft Comput.* 121, 108729. doi: 10.1016/j.asoc.2022.108729
- Hong, J., Zhang, Y.-D., and Chen, W. (2022b). Source-free unsupervised domain adaptation for cross-modality abdominal multi-organ segmentation. *Knowledge Based Syst.* 250, 109155. doi: 10.1016/j.knsys.2022.109155
- Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., and Shen, Y. (2021). Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. *IEEE Trans. Med. Imaging* 41, 145–157. doi: 10.1109/TMI.2021.3107013
- Hu, S., Shen, Y., Wang, S., and Lei, B. (2020a). "Brain MR to PET synthesis via bidirectional generative adversarial network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer), 698–707. doi: 10.1007/978-3-030-59713-9_67
- Hu, S., Yu, W., Chen, Z., and Wang, S. (2020b). "Medical image reconstruction using generative adversarial network for Alzheimer disease assessment with class-imbalance problem," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)* (IEEE), 1323–1327. doi: 10.1109/ICCC51575.2020.9344912
- Hu, S., Yuan, J., and Wang, S. (2019). "Cross-modality synthesis from MRI to PET using adversarial U-Net with different normalization," in *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)* (Shenzhen: IEEE), 1–5. doi: 10.1109/ICMIPE47306.2019.9098219
- Jack, C. R. Jr., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., et al. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* 12, 207–216. doi: 10.1016/S1474-4422(12)70291-0
- Jiang, Y., Chang, S., and Wang, Z. (2021). "TransGAN: two pure transformers can make one strong GAN, and that can scale up," in *Advances in Neural Information Processing Systems* (Montreal: MIT Press). 34, 14745–14758.
- Kabbara, A., Eid, H., El Falou, W., Khalil, M., Wendling, F., and Hassan, M. (2018). Reduced integration and improved segregation of functional brain networks in Alzheimer's disease. *J. Neural Eng.* 15, 026023. doi: 10.1088/1741-2552/aaa76
- Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*. doi: 10.48550/arXiv.1609.02907
- Lei, B., Huang, S., Li, R., Bian, C., Li, H., Chou, Y.-H., et al. (2018). Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder-decoder network. *Neurocomputing* 321, 178–186. doi: 10.1016/j.neucom.2018.09.043

International Partnership Program of Chinese Academy of Sciences under grant GJHZ2021132, and NSFC 62003331.

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.

- Lei, B., Liang, E., Yang, M., Yang, P., Zhou, F., Tan, E.-L., et al. (2022). Predicting clinical scores for Alzheimer's disease based on joint and deep learning. *Expert Syst. Appl.* 187, 115966. doi: 10.1016/j.eswa.2021.115966
- Li, X., Steffens, D. C., Potter, G. G., Guo, H., Song, S., and Wang, L. (2017). Decreased between-hemisphere connectivity strength and network efficiency in geriatric depression. *Hum. Brain Mapp.* 38, 53–67. doi: 10.1002/hbm.23343
- Mo, L.-F., and Wang, S.-Q. (2009). A variational approach to nonlinear two-point boundary value problems. *Nonlinear Anal. Theory Methods Appl.* 71, e834–e838. doi: 10.1016/j.na.2008.12.006
- Mormino, E. C., Papp, K. V., Rentz, D. M., Schultz, A. P., LaPoint, M., Amariglio, R., et al. (2016). Heterogeneity in suspected non-Alzheimer disease pathophysiology among clinically normal older individuals. *JAMA Neurol.* 73, 1185–1191. doi: 10.1001/jamaneurol.2016.2237
- Qiao, L., Zhang, H., Kim, M., Teng, S., Zhang, L., and Shen, D. (2016). Estimating functional brain networks by incorporating a modularity prior. *Neuroimage* 141, 399–407. doi: 10.1016/j.neuroimage.2016.07.058
- Skouras, S., Falcon, C., Tucholka, A., Rami, L., Sanchez-Valle, R., Lladó, A., et al. (2019). Mechanisms of functional compensation, delineated by eigenvector centrality mapping, across the pathophysiological continuum of Alzheimer's disease. *Neuroimage Clin.* 22, 101777. doi: 10.1016/j.nicl.2019.101777
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12, e1001779. doi: 10.1371/journal.pmed.1001779
- Suthaharan, S. (2016). "Support vector machine," in *Machine Learning Models and Algorithms for Big Data Classification*, eds R. Sharda and S. Voß (Boston, MA: Springer), 207–235. doi: 10.1007/978-1-4899-7641-3_9
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Wang, J., Wang, X., Xia, M., Liao, X., Evans, A., and He, Y. (2015). GREYNA: a graph theoretical network analysis toolbox for imaging connectomics. *Front. Hum. Neurosci.* 9, 386. doi: 10.3389/fnhum.2015.00386
- Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., et al. (2007). Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. *Hum. Brain Mapp.* 28, 967–978. doi: 10.1002/hbm.20324
- Wang, S., Chen, Z., You, S., Wang, B., Shen, Y., and Lei, B. (2022). Brain stroke lesion segmentation using consistent perception generative adversarial network. *Neural Comput. Appl.* 34, 8657–8669. doi: 10.1007/s00521-021-06816-8
- Wang, S., Shen, Y., Chen, W., Xiao, T., and Hu, J. (2017). "Automatic recognition of mild cognitive impairment from MRI images using expedited convolutional neural networks," in *International Conference on Artificial Neural Networks* (Alghero: Springer), 373–380. doi: 10.1007/978-3-319-68600-4_43
- Wang, S., Hu, Y., Shen, Y., and Li, H. (2018a). Classification of diffusion tensor metrics for the diagnosis of a myelopathic cord using machine learning. *Int. J. Neural Syst.* 28, 1750036. doi: 10.1142/S0129065717500368
- Wang, S., Shen, Y., Zeng, D., and Hu, Y. (2018b). "Bone age assessment using convolutional neural networks," in *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)* (Chengdu: IEEE), 175–178. doi: 10.1109/ICAIBD.2018.8396189
- Wang, S., Wang, H., Shen, Y., and Wang, X. (2018c). "Automatic recognition of mild cognitive impairment and Alzheimer's disease using ensemble based 3D densely connected convolutional networks," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, eds M. Wani, M. Kantardzic, M. Sayed-Mouchaweh, (Orlando, FL: IEEE), 517–523. doi: 10.1109/ICMLA.2018.00083
- Wang, S., Wang, H., Cheung, A. C., Shen, Y., and Gan, M. (2020a). "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease," in *Deep Learning Applications* (Singapore: Springer), 53–73. doi: 10.1007/978-981-15-1816-4_4
- Wang, S., Wang, X., Hu, Y., Shen, Y., Yang, Z., Gan, M., et al. (2020b). Diabetic retinopathy diagnosis using multichannel generative adversarial network with semisupervision. *IEEE Trans. Automat. Sci. Eng.* 18, 574–585. doi: 10.1109/TASE.2020.2981637
- Wang, S., Wang, X., Shen, Y., He, B., Zhao, X., Cheung, P. W.-H., et al. (2020c). An ensemble-based densely-connected deep learning system for assessment of skeletal maturity. *IEEE Trans. Syst. Man Cybernet. Syst.* 52, 426–437. doi: 10.1109/TSMC.2020.2997852
- Wang, S.-Q. (2009). A variational approach to nonlinear two-point boundary value problems. *Comput. Math. Appl.* 58, 2452–2455. doi: 10.1016/j.camwa.2009.03.050
- Wang, S.-Q., Li, X., Cui, J.-L., Li, H.-X., Luk, K. D., and Hu, Y. (2015). Prediction of myelopathic level in cervical spondylotic myelopathy using diffusion tensor imaging. *J. Magnet. Reson. Imaging* 41, 1682–1688. doi: 10.1002/jmri.24709
- Wee, C.-Y., Yang, S., Yap, P.-T., and Shen, D. (2016). Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. *Brain Imaging Behav.* 10, 342–356. doi: 10.1007/s11682-015-9408-2
- Whitwell, J. L., Przybelski, S. A., Weigand, S. D., Knopman, D. S., Boeve, B. F., Petersen, R. C., et al. (2007). 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain* 130, 1777–1786. doi: 10.1093/brain/awm112
- Xia, M., Wang, J., and He, Y. (2013). BrainNet viewer: a network visualization tool for human brain connectomics. *PLoS ONE* 8, e68910. doi: 10.1371/journal.pone.0068910
- Yang, P., Zhao, C., Yang, Q., Wei, Z., Xiao, X., Shen, L., et al. (2022). Diagnosis of obsessive-compulsive disorder via spatial similarity-aware learning and fused deep polynomial network. *Med. Image Anal.* 75, 102244. doi: 10.1016/j.media.2021.102244
- Ye, C., Mori, S., Chan, P., and Ma, T. (2019). Connectome-wide network analysis of white matter connectivity in Alzheimer's disease. *Neuroimage Clin.* 22, 101690. doi: 10.1016/j.nicl.2019.101690
- You, S., Lei, B., Wang, S., Chui, C. K., Cheung, A. C., Liu, Y., et al. (2022). Fine perceptive GANs for brain MR image super-resolution in wavelet domain. *IEEE Trans. Neural Netw. Learn. Syst.* 1–13. doi: 10.1109/TNNLS.2022.3153088
- Yu, M., Engels, M. M., Hillebrand, A., Van Straaten, E. C., Gouw, A. A., Teunissen, C., et al. (2017). Selective impairment of hippocampus and posterior hub areas in Alzheimer's disease: an MEG-based multiplex network study. *Brain* 140, 1466–1485. doi: 10.1093/brain/aww050
- Yu, S., Wang, S., Xiao, X., Cao, J., Yue, G., Liu, D., et al. (2020). "Multi-scale enhanced graph convolutional network for early mild cognitive impairment detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Lima: Springer), 228–237. doi: 10.1007/978-3-030-59728-3_23
- Yu, W., Lei, B., Ng, M. K., Cheung, A. C., Shen, Y., and Wang, S. (2021). Tensorizing GAN with high-order pooling for Alzheimer's disease assessment. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 4945–4959. doi: 10.1109/TNNLS.2021.3063516
- Yu, W., Lei, B., Wang, S., Liu, Y., Feng, Z., Hu, Y., et al. (2022). Morphological feature visualization of Alzheimer's disease via multidirectional perception GAN. *IEEE Trans. Neural Netw. Learn. Syst.* 1–15. doi: 10.1109/TNNLS.2021.3118369
- Yuzwa, S. A., Macauley, M. S., Heinonen, J. E., Shan, X., Dennis, R. J., He, Y., et al. (2008). A potent mechanism-inspired o-glucanase inhibitor that blocks phosphorylation of tau in vivo. *Nat. Chem. Biol.* 4, 483–490. doi: 10.1038/nchembio.96
- Zeng, D., Wang, S., Shen, Y., and Shi, C. (2017). A GA-based feature selection and parameter optimization for support tucker machine. *Proc. Comput. Sci.* 111, 17–23. doi: 10.1016/j.procs.2017.06.004
- Zhang, H., Wang, Y., Wang, Y., Li, X., Wang, S., and Wang, Z. (2022). Recent advance on carbamate-based cholinesterase inhibitors as potential multifunctional agents against Alzheimer's disease. *Eur. J. Med. Chem.* 240, 114606. doi: 10.1016/j.ejmech.2022.114606
- Zhou, Y., Zhang, L., Teng, S., Qiao, L., and Shen, D. (2018). Improving sparsity and modularity of high-order functional connectivity networks for mci and asd identification. *Front. Neurosci.* 12, 959. doi: 10.3389/fnins.2018.00959
- Zuo, Q., Lei, B., Shen, Y., Liu, Y., Feng, Z., and Wang, S. (2021a). "Multimodal representations learning and adversarial hypergraph fusion for early Alzheimer's disease prediction," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)* (Zhuhai: Springer), 479–490. doi: 10.1007/978-3-030-88010-1_40
- Zuo, Q., Lei, B., Wang, S., Liu, Y., Wang, B., and Shen, Y. (2021b). A prior guided adversarial representation learning and hypergraph perceptual network for predicting abnormal connections of Alzheimer's disease. *arXiv preprint arXiv:2110.09302*. doi: 10.48550/arXiv.2110.09302



OPEN ACCESS

EDITED BY

Zhiguo Zhang,
Harbin Institute of Technology, China

REVIEWED BY

Yanghua Tian,
The First Affiliated Hospital of Anhui
Medical University, China
Peng Li,
Shenzhen University, China

*CORRESPONDENCE

Li Gao
gaoli@shu.edu.cn
Xiaosong He
hexs@ustc.edu.cn
Xiaochu Zhang
zxcustc@ustc.edu.cn

†These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

RECEIVED 12 November 2022

ACCEPTED 24 November 2022

PUBLISHED 09 December 2022

CITATION

Chen Y, Wei Z, Gou H, Liu H, Gao L,
He X and Zhang X (2022) How far is
brain-inspired artificial intelligence
away from brain?
Front. Neurosci. 16:1096737.
doi: 10.3389/fnins.2022.1096737

COPYRIGHT

© 2022 Chen, Wei, Gou, Liu, Gao, He
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.

How far is brain-inspired artificial intelligence away from brain?

Yucan Chen^{1†}, Zhengde Wei^{2†}, Huixing Gou³, Haiyi Liu⁴,
Li Gao^{5*}, Xiaosong He^{2*} and Xiaochu Zhang^{1,2,6,7*}

¹Hefei National Research Center for Physical Sciences at the Microscale, and Department of Radiology, the First Affiliated Hospital of USTC, Division of Life Science and Medicine, University of Science & Technology of China, Hefei, China, ²Department of Psychology, School of Humanities and Social Sciences, University of Science and Technology of China, Hefei, Anhui, China, ³Division of Life Sciences and Medicine, School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, China, ⁴State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China, ⁵SILC Business School, Shanghai University, Shanghai, China, ⁶Application Technology Center of Physical Therapy to Brain Disorders, Institute of Advanced Technology, University of Science and Technology of China, Hefei, China, ⁷Biomedical Sciences and Health Laboratory of Anhui Province, University of Science and Technology of China, Hefei, China

Fueled by the development of neuroscience and artificial intelligence (AI), recent advances in the brain-inspired AI have manifested a tipping-point in the collaboration of the two fields. AI began with the inspiration of neuroscience, but has evolved to achieve a remarkable performance with little dependence upon neuroscience. However, in a recent collaboration, research into neurobiological explainability of AI models found that these highly accurate models may resemble the neurobiological representation of the same computational processes in the brain, although these models have been developed in the absence of such neuroscientific references. In this perspective, we review the cooperation and separation between neuroscience and AI, and emphasize on the current advance, that is, a new cooperation, the neurobiological explainability of AI. Under the intertwined development of the two fields, we propose a practical framework to evaluate the brain-likeness of AI models, paving the way for their further improvements.

KEYWORDS

artificial intelligence, brain, brain-inspired intelligence, neurobiological explainability, AI evaluation, artificial neural network

Introduction

Artificial intelligence (AI) starts with the notion of creating Turing-powerful intelligent systems (Turing, 1936). He claimed that his desire was to build a machine to “imitate a brain” and also to “mimic the behavior of the human,” which means the likeness to both the brain and the behavior is requisite to realize such intelligent systems. For this to happen, pioneers in the field (Rosenblatt, 1958; Fukushima and Nixon, 1980; Bi and Poo, 1998; Masquelier and Thorpe, 2007) have drawn inspiration

from the neurobiological representation to develop AI models. However, early models or algorithms strictly mimicking the neural processes in the brain have constantly failed to deliver satisfactory performances, such as the perceptron (Rosenblatt, 1958), Hebbian learning rules (Kempster et al., 1999), and Sigmoid (Han and Moraga, 1995). Gradually, computer scientists have strayed away from neuroscience and turned to engineering and mathematical solutions to design “outcome-driven” models. These models achieved remarkable performance in many aspects, including but not limited to object recognition (Riesenhuber and Poggio, 2000), speech and music recognition (Kell et al., 2018; Sutskever et al., 2019), and motor movement (Todorov, 2000).

Nonetheless, comparison between AI and the brain has never stopped. Once optimized performance is achieved, researchers (Yamins et al., 2014; Güçlü and van Gerven, 2015; Eickenberg et al., 2017; Zhuang et al., 2021) begin to search for the neurobiological explainability of these advanced models, that is, the similarity of the neurobiological representation of the same computational processes between AI models and the brain. The authors wish that through unraveling the neurobiological explainability of AI models, one could achieve a better understanding of the brain and thus promote the development of neuroscience (Lindsay, 2021). Interestingly, in return, the evaluation of resemblance between current AI models and the brain may also shed lights on how far away these models are to the Turing-powerful (i.e., brain-like) intelligent systems.

During the three stages (Figure 1) of AI development, the role of neuroscience has experienced a shift from the “guide,” who provides guiding principles to the design of AI models, to the “judge,” who provide references for the evaluation of AI models. In this review, we will look back to the mutual development of AI and neuroscience, and propose a framework to evaluate the brain-likeness of AI models that can serve AI development in multiple ways.

The collaboration and separation of artificial intelligence and neuroscience

Brain is the most complex and efficient non-artificial intelligent system known to humans. Throughout history, the promise of creating machine intelligence with brain-like ability has been a motivation of innovation (Roy et al., 2019). One way to realize such intelligence is to scrutinize the organization principles of brain's structures and functions and thus seek inspiration for the design of AI. Hassabis et al. (2017) stated that if a new facet of neurobiological representation were found, it would be considered as an excellent candidate for incorporation into AI. Over the years, AI models have been rapidly developed by drawing

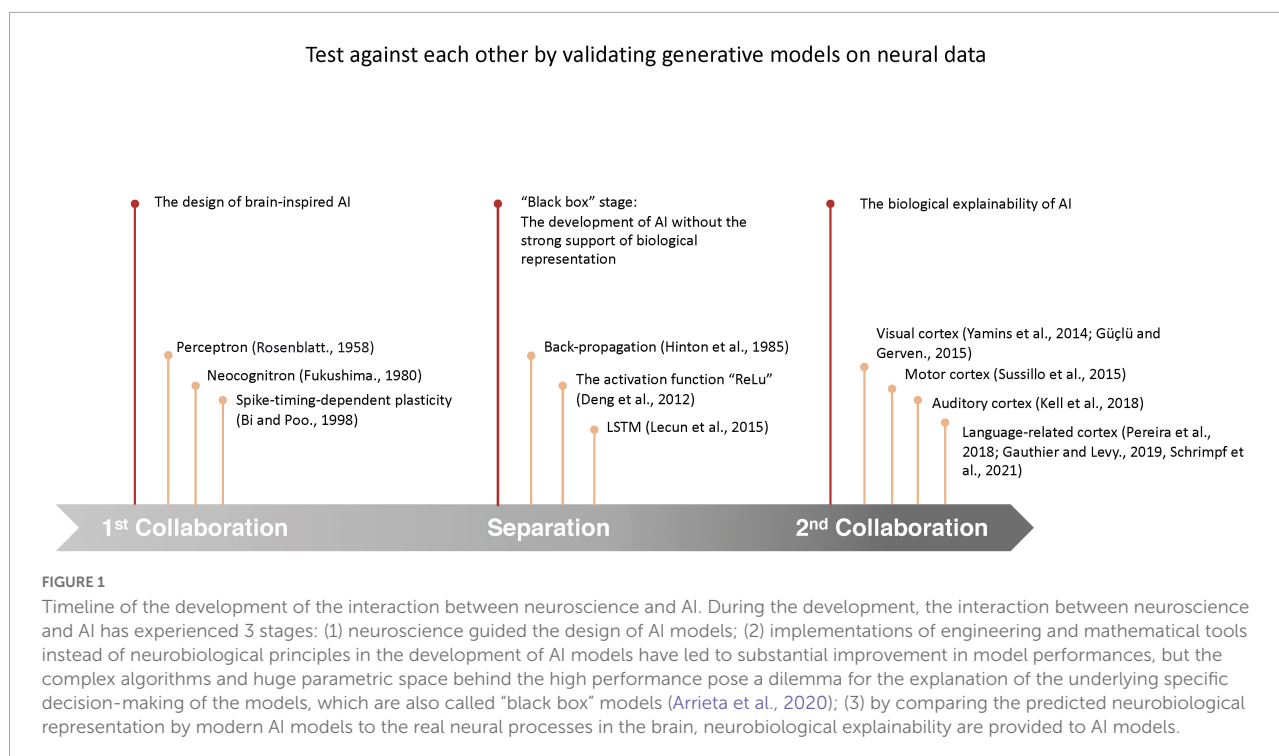
inspiration from the brain neural networks, whereas algorithms, architectures, and functions of models have benefited greatly from mimicking such neurobiological representations (e.g., neuro-synaptic framework and hierarchical structure).

In the initial collaboration between AI and neuroscience, the direct inspiration from neuroscience accelerated the start-up of AI. The earliest application was the perceptron (Rosenblatt, 1958), a simple abstract of neurons, mimicking the simple neuronal activity in visual cortex, such as the weights of synapses, the biases of the thresholds, and the activation function of the neural cells. Years later, inspired by Hubel and Wiesel's (1962) study in the visual cortex, Fukushima and Nixon (1980) proposed an advanced model, Neocognitron, the precursor of the modern convolutional neural networks (CNN), which mimicked the organizations of neural cells in the visual cortex. Apart from the inspiration of how neurons activate, researchers also designed some brain-corresponding models (e.g., topographic maps) inspired by how brain is organized. For example, Burak and Fiete (2009) modeled the network topology of the rats entorhinal cortex to form the neural substrate for dead-reckoning.

Although AI is profoundly inspired by the neurobiological representation of the brain, surprisingly, these brain-mimicking models have never achieved a satisfactory performance, likely due to their over-simplification of the real neural system. For instance, Hebbian learning, a neurobiologically schemed method, fail to produce models with adequate performance as it does not take into consideration of the synapse's downstream effect on the network output (Lillicrap et al., 2020). Gradually, researchers (Rumelhart et al., 1986; Hinton et al., 2012; Lecun et al., 2015) started to turn to engineering and mathematical solutions to maximize model performance regardless of its underlying neurobiological relevance. In these works, the authors replaced the former neurobiological schemed methods with back-propagation, an algorithm without a prior neurobiological relevance, and solved the low-efficiency problem of synaptic modification (Lillicrap et al., 2020). Moreover, replacing the former neurobiologically inspired Sigmoid function (Han and Moraga, 1995) with the activation function ReLu (Deng et al., 2010) has been demonstrated to substantially improve the performance of deep neural networks (DNNs) since Krizhevsky et al. (2012). Given such superior performances, are these models operated in anyway similar to the most efficient system we ever know, the brain, despite they are not strictly structured to follow any neurobiological principles?

Neurobiological explainability of artificial intelligence

Despite of the turning of design principles from mimicking neurobiological representation of the brain to optimizing



performance with tools from engineering and mathematics, AI and neuroscience have never really grown apart. With the rapid development of AI, researchers (Yamins et al., 2014; Güçlü and van Gerven, 2015; Eickenberg et al., 2017; Zhuang et al., 2021) believe that these advanced models are capable to promote the development of neuroscience in return. In specific, they advocate for seeking for the neurobiological explanations for AI models as an alternative way to better understand the organization principles of the brain.

Early studies exploring the neurobiological explainability of AI models have mainly focused on visual recognition. Yamins et al. (2014) first examined the similarity between real brain activities and predicted activations from CNN model. The authors trained CNN model to match human performances on various visual recognition tasks. The results showed that the third and the fourth (top) layer of the model could effectively predict the inferior temporal activity recorded with functional MRI during image recognition. Other findings (Cadieu et al., 2014; Khaligh-Razavi and Kriegeskorte, 2014; Güçlü and van Gerven, 2015) also confirmed that deep neural network (DNN) models trained for visual recognition have remarkable predictability for the neural responses in the human visual system as well. Moreover, Cichy et al. (2016) found that the predicted brain activities by DNN trained for object categorization are highly resemble to the brain activations recorded *via* both fMRI and MEG during the same cognitive process, not only in the physical space domain (i.e., matching the hierarchical topography in the human ventral and dorsal

visual streams), but also in the temporal domain (i.e., matching the time course over visual processing).

In addition to visual recognition, models designed for other utilities also showed similar predicted neurobiological representations with the real activities in the corresponding neural systems. A recent heavily focused area is the neurobiological explainability of AI models for language processing, including syntax processing (Gauthier and Levy, 2019), semantic processing (Pereira et al., 2016; de Heer et al., 2017), and comprehension (Schrimpf et al., 2021). Adding to these evidence, highly corresponded mappings between predicted (by the AI) and recorded (in the brain) neurobiological representations have also been found in other cognitive systems, such as the auditory system (Kell et al., 2018), the motor system (Sussillo et al., 2015), and even the hippocampal formation (Whittington et al., 2021).

Quantify the progress toward Turing-powerful intelligence

Such demonstration of neurobiological explainability of AI models has opened the door for new contributions from neuroscience, to provide alternative tools to quantitatively evaluate the progress we made toward the Turing-powerful intelligent systems. Normally, evaluation to the distance to such intelligent systems concentrates at the behavioral level, where the performance of models would be evaluated, such as model-model comparison and model-to-human behavior comparison.

However, the neurobiological explainability gives us cues to evaluate the brain-likeness of the models, which mainly focused on whether they can solve the same problems as the brain. Evaluation at the both behavioral and neurobiological gives us a more comprehensive insight to evaluate the distance to the Turing-powerful intelligent systems. Besides, the improvements of the algorithm can also indicate the advancement toward the Turing-powerful intelligent systems. To further elaborate on this new role of neuroscience in AI development, here, we capitalize on the Marr's (1982) widely recognized computational framework, and discuss such applications in three levels.

Evaluation at the computational level

In Marr's theory, the first level, computational level, concerns the problems that models can solve. The evaluation of the performance of the models can be categorized into two ways: model-to-model comparison and model-to-human behavior comparison. The model-to-model comparison literally compares performances of different models for the same task. For instance, Xu et al. (2021) compared supervised models to unsupervised models and found that the latter trained with 10 min of labeled data, could rival the best supervised model trained with 960 h of labeled data. The model-to-human behavior comparison contrasts AI performance to human performance during the same task. For instance, Rajalingham et al. (2018) compared the ANNs' (Artificial neural networks) accuracy in the visual categorization task with the behavioral results from 1,477 primates (1,472 humans and 5 monkeys), and evaluated that the models could not achieve the human-like behavioral performance.

Evaluation at the algorithmic level

The second level of Marr's theory, algorithmic level, concerns the processes that models go through. During the exploration into the neurobiological explainability of models, the training methods for models displayed a positive shift, implying that models turned out to be more intelligent. First, the training methods for models in the earlier studies aimed to map the computational models into the corresponding brain activity (Mitchell et al., 2008) when receiving the same stimuli, or to use the brain responses to constrain the models (Cadieu et al., 2014). And then the artificial neural response generated from models and the unlearned brain data were compared. However, in more recent studies, researchers (Yamins et al., 2014; Kell et al., 2018; Schrimpf et al., 2021) start to train AI models with only behavioral (e.g., objects and their labels) but not any neuroimaging data. Interestingly, while the models were not optimized to fit brain signals in the first places, they can still predict the brain responses during the same cognitive process

proficiently. These findings suggest that the computational processes of these models can be brain-like enough to generate neurobiological representation without explicit training.

Furthermore, the shift of paradigm from supervised to unsupervised models during the prediction of neurobiological representation can also be seen as a step-forward toward brain-like intelligence, since the latter is considered to be more similar to human learning pattern which is constantly exposed to unlabeled environments (Mitchell, 2004), which could even automatically learn the human bias from image classification (Steed and Caliskan, 2021). In earlier studies, models used to predict neurobiological representation were mostly supervised models (Cadieu et al., 2014; Yamins et al., 2014; Güçlü and van Gerven, 2015). A study even suggested that unsupervised models could not predict the brain responses (Khaligh-Razavi and Kriegeskorte, 2014). However, with the improvement of unsupervised models during the decade (Xu et al., 2021), recent studies have found evidence that unsupervised models could successfully predict the neural response as well. For instance, Zhuang et al. (2021) found that the unsupervised models achieved a high prediction accuracy in the primate ventral stream that equaled and even surpassed the performance of the best supervised model. Thus, the recent success of unsupervised models in predicting brain representation suggests that AI models have made a giant step forward on the human-like path.

Evaluation at the implementation/physical level

The last level of Marr's theory, implementation/physical level, concerns the brain-likeness of the models.

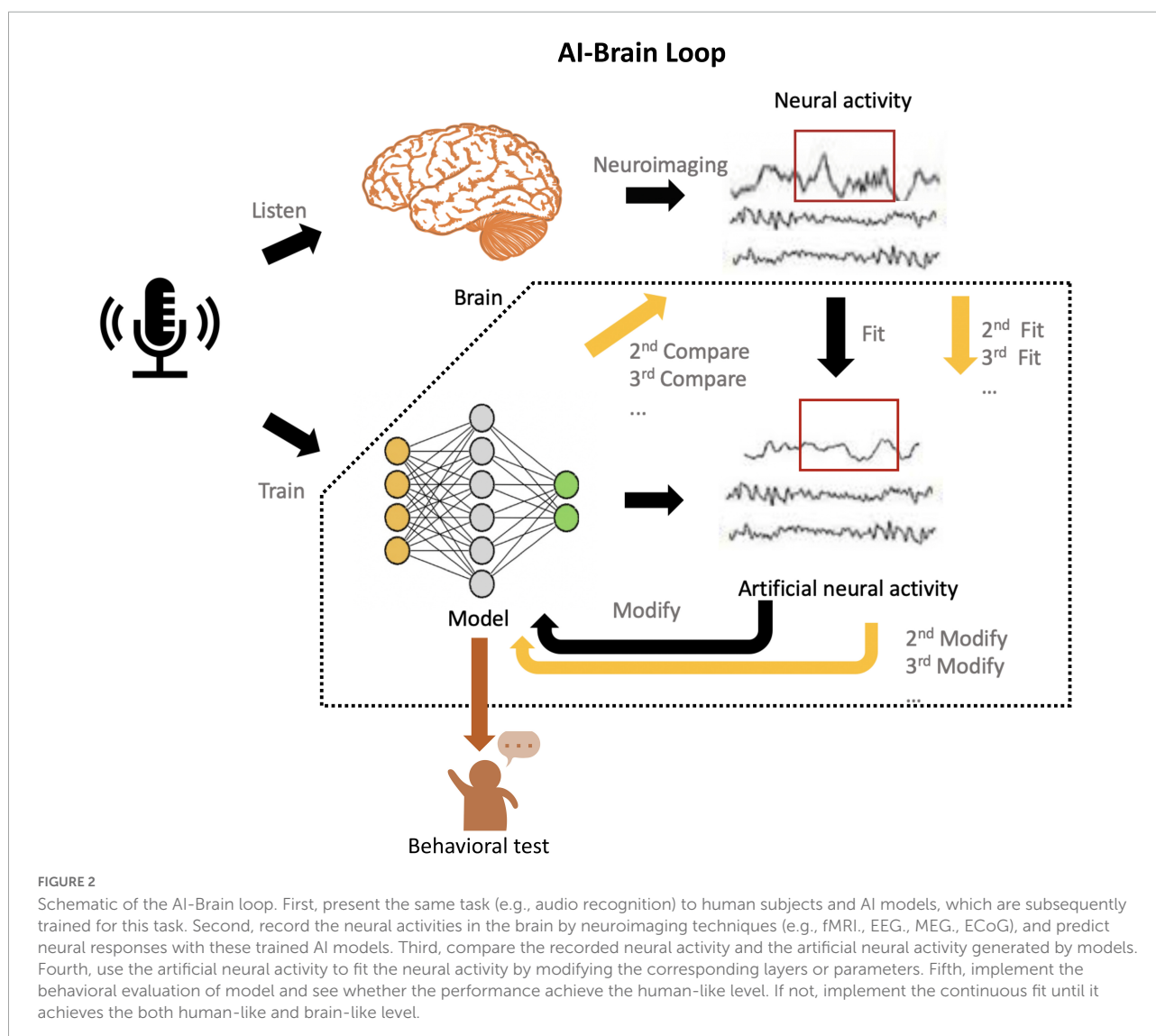
First, instantiation (i.e., the neural representation of models) of the brain-inspired model would be an explicit and effective measure for judging success and spurring the progress to the Turing-powerful intelligence. It is an explicit measure since the layers in the models almost correspond to the hierarchical structure of the brain (Kell et al., 2018), where we can directly compared the detailed performance in each layer with the corresponding responses in the brain. If a model could highly predict the response in the brain, we consider that its corresponding parameters/weight would be vital to achieve the Turing-powerful intelligent systems (Hassabis et al., 2017). It is also an effective measure to drive models toward the goal. For example, Yamins et al. (2014) found that the top layer in the model could better predict the activity in IT cortex while other layers did not achieve the satisfactory performance. In this case, we may allocate more energy to optimize the layers that cannot successfully explain the corresponding neurobiological representation, which determines the most productive way to allocate resources.

Second, the evaluation of models from the perspective of neuroscience further supports the validation of behavioral

results. Many studies have indicated that the more brain-like the model is, the better performance the model has in the task. For instance, [Yamins et al. \(2014\)](#) indicated that when a model highly predicts the IT (Inferior temporal) cortex, the better performance it would have in the object recognition task. And the same results were also found in another study ([Khaligh-Razavi and Kriegeskorte, 2014](#)). They compared 37 models with the human's and monkeys' cortex, respectively, showing that the models with more relevant correspondences with the neural representation in IT cortex have better performance in object recognition. Further, in the language model, [Schrimpf et al. \(2021\)](#) compared the predictability for neural response between 43 diverse language models, where they found models with high next-word predictive ability, like GPT models, have a better performance in predicting brain signals in language comprehension. Even they compared the ability of the next-word prediction of these models in another dataset, the neural

predictivity still significantly correlated with the behavioral results. The parallel but highly correlated results provide us an opportunity to evaluate and further modify models from another perspective, neuroscience. Combining with the first point, it gives us a sight that we may modify the models more brain-like in order to achieve better performance.

Third, the evaluation at the behavioral level may not comprehensively explain the brain-like intelligence, as the way to process information differs in the brain and behavior ([Bechara et al., 1997](#); [Soon et al., 2008](#)). Researchers claimed that the unconscious biases observed in the brain guided behavior before the conscious knowledge did, which means the brain signal might capture the subtle differences that were obscure at the behavioral level. Thus, the evaluation at the neurobiological level may evaluate the distance to the Turing-powerful intelligence more accurately compared to the behavioral evaluation.



Implications for the improvement of AI models

Thus far, we have reviewed the collaboration and separation between neuroscience and AI, and highlighted the significance of the current collaboration. More importantly, we propose the importance of evaluating models from the perspective of neuroscience. The evaluation tells us the closeness between the current models and the brain, which is critical to optimize models in achieving the Turing-powerful level.

To move forward, here, we present an AI-brain loop framework in which we implement the explicit evaluation from neuroscience and accurate modification in each layer and even parameters to the models (Figure 2), inspired by the human-in-the-loop (Li et al., 2014) and inception loop (Walker et al., 2019).

In this framework, we propose that the AI models trained for specific behavioral task can use neural recordings during the same task as neurobiological reference. Comparisons between recorded and model-predicted neural responses can be used to tune the parameter space of the AI models, and more realistic neurobiological representation of the models can be achieved during the process of minimizing the differences between the two. Lastly, performance of the modified models at behavioral level will be used to verify that whether models with higher brain-likeness level, but also function at the human-like level. Then we also test the modified models at the behavioral level and see whether the performance improves. Such clues of the modification are fundamental to achieve the Turing-powerful intelligent system since it echoes Turing's claim (Turing, 1936) that models are qualified in not only "mimicking the behavior of the human," but also "imitating the brain."

The call for Turing-powerful intelligent system asks to look beyond performance optimization, but to focus more on how to achieve higher brain resemblance in future AI models. We believe that re-introducing neuroscience back into AI development through this neurobiological explainability provides a promising opportunity to outbreaking the "black-box" dilemma suffered by most of modern AI models. By "jumping out of the box" and developing more brain-like AI through such AI-brain comparisons, we may eventually leap forward to such ultimate goal.

Author contributions

YC, ZW, XH, and XZ contributed to the conception of the study. YC and ZW wrote the first draft of the manuscript. YC, ZW, and XH wrote sections of the manuscript. YC, ZW, HG, HL, LG, XH, and XZ contributed to revise the manuscript and customized the tables. All authors contributed to manuscript revision, read, and approved the submitted final versions.

Funding

This work was supported by grants from The Chinese National Programs for Brain Science and Brain-like Intelligence Technology (2021ZD0202101), The National Natural Science Foundation of China (71942003, 32161143022, 32171080, and 31900766), Major Project of Philosophy and Social Science Research, Ministry of Education of China (19JZD010), CAS-VPST Silk Road Science Fund 2021 (GLHZ202128), Collaborative Innovation Program of Hefei Science Center, CAS (2020HSC-CIP001), and Anhui Provincial Key Research and Development Project (202004b11020013).

Acknowledgments

We thank the Bioinformatics Center of the University of Science and Technology of China, School of Life Science, for providing supercomputing resources for this project.

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

- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* 58, 82–115. doi: 10.1016/j.inffus.2019.12.012
- Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science* 275, 1293–1295. doi: 10.1126/science.275.5304.1293
- Bi, G. Q., and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472. doi: 10.1523/JNEUROSCI.18-24-10464.1998
- Burak, Y., and Fiete, I. R. (2009). Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* 5:e1000291. doi: 10.1371/journal.pcbi.1000291
- Cadiou, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., et al. (2014). Deep neural networks reveal the representation of primate IT cortex for core visual object recognition. *PLoS Comput. Biol.* 10:e1003963. doi: 10.1371/JOURNAL.PCBI.1003963
- Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* 6:27755. doi: 10.1038/srep27755
- de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., and Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *J. Neurosci.* 37, 6539–6557. doi: 10.1523/JNEUROSCI.3267-16.2017
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2010). “ImageNet: A large-scale hierarchical image database,” in *Proceedings of the 2009 IEEE conference on computer vision and pattern recognition*, (Miami, FL: IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848
- Eickenberg, M., Gramfort, A., Varoquaux, G., and Thirion, B. (2017). Seeing it all: Convolutional network layers map the function of the human visual system. *Neuroimage* 152, 184–194. doi: 10.1016/j.neuroimage.2016.10.001
- Fukushima, T., and Nixon, J. C. (1980). Analysis of reduced forms of bioprotein in biological tissues and fluids. *Anal. Biochem.* 102, 176–188. doi: 10.1016/0003-2697(80)90336-X
- Gauthier, J., and Levy, R. P. (2019). “Linking artificial and human neural representations of language,” in *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, (Hong Kong: Association for Computational Linguistics), 529–539. doi: 10.18653/V1/D19-1050
- Güçlü, U., and van Gerven, M. A. J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* 35, 10005–10014. doi: 10.1523/JNEUROSCI.5023-14.2015
- Han, J., and Moraga, C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. *Lect. Notes Comput. Sci.* 930, 195–201. doi: 10.1007/3-540-59497-3_175/COVER
- Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron* 95, 245–258. doi: 10.1016/j.neuron.2017.06.011
- Hinton, G. E., Srastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arX [Preprint]* doi: 10.48550/arXiv.1207.0580
- Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/JPHYSIOL.1962.SP006837
- Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V., and McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron* 98, 630–644.e16. doi: 10.1016/j.neuron.2018.03.044
- Kempler, R., Gerstner, W., and van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Phys. Rev. E* 59:4498. doi: 10.1103/PhysRevE.59.4498
- Khaligh-Razavi, S. M., and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* 10:e1003915. doi: 10.1371/JOURNAL.PCBI.1003915
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105.
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Li, W., Sadigh, D., Sastry, S. S., and Seshia, S. A. (2014). Synthesis for human-in-the-loop control systems. *Lect. Notes Comput. Sci.* 8413, 470–484. doi: 10.1007/978-3-642-54862-8_40/COVER
- Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., and Hinton, G. (2020). Backpropagation and the brain. *Nat. Rev. Neurosci.* 21, 335–346. doi: 10.1038/s41583-020-0277-3
- Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: Past, present, and future. *J. Cogn. Neurosci.* 33, 2017–2031. doi: 10.1162/JOCN_A_01544
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: W.H. Freeman and Company.
- Masquelier, T., and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.* 3:e31. doi: 10.1371/journal.pcbi.0030031
- Mitchell, T. M. (2004). “The role of unlabeled data in supervised learning,” in *Language, knowledge, and representation*, Vol. 99, eds J. M. Larrazabal and L. A. P. Miranda (Dordrecht: Springer), 103–111. doi: 10.1007/978-1-4020-2783-3_7
- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science* 320, 1191–1195. doi: 10.1126/science.1152876
- Pereira, F., Gershman, S., Ritter, S., and Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cogn. Neuropsychol.* 33, 175–190. doi: 10.1080/02643294.2016.1176907
- Rajalingham, R., Issa, E. B., Bashan, P., Kar, K., Schmidt, K., and DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *J. Neurosci.* 38, 7255–7269. doi: 10.1523/JNEUROSCI.0388-18.2018
- Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. *Nat. Neurosci.* 3, 1199–1204. doi: 10.1038/81479
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65, 386–408. doi: 10.1037/H0042519
- Roy, K., Jaiswal, A., and Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature* 575, 607–617. doi: 10.1038/s41586-019-1677-2
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536. doi: 10.1038/323533a0
- Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., et al. (2021). The neural architecture of language: Integrate modeling converges on predictive processing. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2105646118. doi: 10.1073/pnas.2105646118
- Soon, C. S., Brass, M., Heinze, H. J., and Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nat. Neurosci.* 11, 543–545. doi: 10.1038/nn.2112
- Steed, R., and Caliskan, A. (2021). “Image representations learned with unsupervised pre-training contain human-like biases,” in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, (Toronto: IEEE), 701–713. doi: 10.1145/3442188.3445932
- Sussillo, D., Churchland, M. M., Kaufman, M. T., and Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* 18, 1025–1033. doi: 10.1038/nn.4042
- Sutskever, I., Martens, J., and Hinton, G. E. (2019). “Generating text with recurrent neural networks,” in *Proceedings of the 28th international conference on machine learning*, Bellevue, WA, 1017–1024.
- Todorov, E. (2000). Direct cortical control of muscle activation in voluntary arm movements: A model. *Nat. Neurosci.* 3, 391–398. doi: 10.1038/73964

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *J. Math.* 58, 345–363.

Walker, E. Y., Sinz, F. H., Cobos, E., Muhammad, T., Froudarakis, E., Fahey, P. G., et al. (2019). Inception loops discover what excites neurons most using deep predictive models. *Nat. Neurosci.* 22, 2060–2065. doi: 10.1038/s41593-019-0517-x

Whittington, J. C. R., Warren, J., and Behrens, T. E. J. (2021). Relating transformers to models and neural representations of the hippocampal formation. *arX [Preprint]* doi: 10.48550/arXiv.2112.04035

Xu, Q., Baevski, A., Likhomanenko, T., Tomasello, P., Conneau, A., Collobert, R., et al. (2021). “Self-training and pre-training are complementary for speech

recognition,” in *Proceedings of the ICASSP, IEEE international conference on acoustics, speech and signal processing*, (Toronto: IEEE), 3030–3034. doi: 10.1109/ICASSP39728.2021.9414641

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8619–8624. doi: 10.1073/pnas.1403112111

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., et al. (2021). Unsupervised neural network models of the ventral visual stream. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2014196118. doi: 10.1073/PNAS.2014196118/SUPPL_FILE/PNAS.2014196118.SAPP.PDF



OPEN ACCESS

EDITED BY

Shuqiang Wang,
Shenzhen Institutes of Advanced Technology
(CAS), China

REVIEWED BY

Weixiang Liu,
Shenzhen University, China
Linglong Qian,
King's College London, United Kingdom

*CORRESPONDENCE

Jiang Zhong
✉ zhongjiang@cqu.edu.cn
Shuai Han
✉ hanshuai19870217@outlook.com

SPECIALTY SECTION

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

RECEIVED 30 November 2022

ACCEPTED 17 January 2023

PUBLISHED 10 February 2023

CITATION

Wang C, Yu J, Zhong J, Han S, Qi Y, Fang B and
Li X (2023) Prior knowledge-based precise
diagnosis of blend sign from head computed
tomography. *Front. Neurosci.* 17:1112355.
doi: 10.3389/fnins.2023.1112355

COPYRIGHT

© 2023 Wang, Yu, Zhong, Han, Qi, Fang 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.

Prior knowledge-based precise diagnosis of blend sign from head computed tomography

Chen Wang¹, Jiefu Yu², Jiang Zhong^{1*}, Shuai Han^{3*}, Yafei Qi⁴,
Bin Fang¹ and Xue Li⁵

¹College of Computer Science, Chongqing University, Chongqing, China, ²Department of Neurosurgery, The First Hospital of China Medical University, Shenyang, China, ³Department of Neurosurgery, Shengjing Hospital of China Medical University, Shenyang, China, ⁴College of Computer Science and Engineering, Central South University, Changsha, China, ⁵School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia

Introduction: Automated diagnosis of intracranial hemorrhage on head computed tomography (CT) plays a decisive role in clinical management. This paper presents a prior knowledge-based precise diagnosis of blend sign network from head CT scans.

Method: We employ the object detection task as an auxiliary task in addition to the classification task, which could incorporate the hemorrhage location as prior knowledge into the detection framework. The auxiliary task could help the model pay more attention to the regions with hemorrhage, which is beneficial for distinguishing the blend sign. Furthermore, we propose a self-knowledge distillation strategy to deal with inaccuracy annotations.

Results: In the experiment, we retrospectively collected 1749 anonymous non-contrast head CT scans from the First Affiliated Hospital of China Medical University. The dataset contains three categories: no intracranial hemorrhage (non-ICH), normal intracranial hemorrhage (normal ICH), and blend sign. The experimental results demonstrate that our method performs better than other methods.

Discussion: Our method has the potential to assist less-experienced head CT interpreters, reduce radiologists' workload, and improve efficiency in natural clinical settings.

KEYWORDS

blend sign, intracranial hemorrhage, hemorrhage expansion, prior knowledge, self-knowledge distillation, convolutional neural network

Introduction

Intracranial hemorrhage (ICH) is a serious neurological disorder. It accounts for about 30% of the whole number of patients with stroke (Qureshi et al., 2009). Many factors such as congenital development, vascular disease, and head injury could lead to ICH (Heit et al., 2017). According to the hemorrhage location, some recent studies (Qureshi et al., 2001) subdivide ICH into intraparenchymal hemorrhage (IPH), intraventricular hemorrhage (IVH), epidural hemorrhage (EDH), subdural hemorrhage (SDH), and subarachnoid hemorrhage (SAH) (Qureshi et al., 2001). Recently, some researchers have paid much attention to the blend sign and black hole sign, two new types of ICH (Li et al., 2015, 2016). Blend sign (Li et al., 2015) is composed of two parts with apparently different CT attenuation. There is a well-defined margin (Li et al., 2017) between the hyperattenuated and relatively hypoattenuated regions, as shown in Figure 1. Some recent studies (Yagi et al., 2019; Zhang et al., 2020; Li et al., 2021; Yang et al., 2021) have shown that blend sign and black hole sign are closely associated with hemorrhage expansion in ICH.

Non-contrast head computed tomography (CT) is a well-known and practical imaging approach for the diagnosis of intracranial hemorrhage (Heit et al., 2017). In the non-contrast CT slices, regions with intracranial hemorrhage appear highlighted since blood has a slightly higher

density than other brain tissues (Nguyen et al., 2016). Patients in emergency departments usually need an evaluation of head CT. In general, the precise diagnosis of intracranial hemorrhage is crucial in patients to assess the need for clinical treatment (Chilamkurthy et al., 2018). Most hospitals usually provide CT scan interpretations by junior radiologists or emergency physicians. Then some senior radiologists will review the initial interpretations. The CT scan interpretation is time-consuming, of low quality, and is unreliable. Several studies have confirmed that some misinterpretations may even lead to clinical consequences (Alfaro et al., 1995; Strub et al., 2007). A precise diagnosis system for intracranial hemorrhage from head CT scans is desirable.

Artificial intelligence and deep learning have recently shown great performance in medically assisted diagnosis (Shin et al., 2016; Havaei et al., 2017; Kamnitsas et al., 2017; Chen et al., 2018). Chen et al. (2018) presented a 3D U-Net to segment cranial vasculature in CTA volume without manual annotations. Havaei et al. (2017) proposed a brain tumor segmentation model with deep neural networks. Kamnitsas et al. (2017) provided an efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Shin et al. (2016) utilized convolutional neural networks for computer-aided detection problems. Recently, some researchers have introduced generative learning into brain disease diagnosis (Wang et al., 2018, 2022; Hu et al., 2019, 2021; Yu et al., 2021, 2022; You et al., 2022). Wang et al. (2018) presented a convolutional neural network-based framework for bone age assessment. Hu et al. (2019) proposed one adversarial U-Net with different normalizations for cross-modality synthesis from MRI to PET. Yu et al. (2021) applied GAN with high-order pooling for Alzheimer's disease. Yu et al. (2022) introduced a novel multi-directional perception generative adversarial network to visualize the morphological features of Alzheimer's disease. Hu et al. (2021) introduced bidirectional mapping generative adversarial networks for brain MRI to PET synthesis. You et al. (2022) designed a fine perceptive generative adversarial network to produce high-resolution MR images from low-resolution counterparts in the wavelet domain. Wang et al. (2022) proposed a segmentation model for brain stroke lesions with consistent perception generative adversarial networks.

Some researchers have explored the detection of abnormalities in head CT with machine learning and deep learning methods (Xiao et al., 2010; Li et al., 2012; Chang et al., 2018; Titano et al., 2018). Classification-based approach is the conventional approach. Li et al. (2012) reported a machine-learning algorithm with high diagnostic value for SAH detection. Prevedello et al. (2017) proposed one small deep-learning model to detect critical test findings for head CT. Chilamkurthy et al. (2018) proposed a deep-learning algorithm for detecting critical findings in head CT scans. They retrospectively collected 4,304 scans for evaluation. Ye et al. (2019) introduced a three-dimensional (3D) joint convolutional neural network for the classification of five subcategories of ICH. Lee et al. (2019) presented an explainable deep-learning algorithm for ICH classification with a small dataset. One disadvantage of these classification-based methods is that the model may fit some unimportant features, such as the background. Moreover, some researchers (Grewal et al., 2018; Liu et al., 2021) introduced the segmentation into the diagnosis of head CT. Grewal et al. (2018) applied three segmentation tasks as auxiliary

tasks to guide the classification model's attention to the regions with hemorrhage. Kuo et al. (2019) proposed an expert-level detection model for acute intracranial hemorrhage, which performed joint classification tasks and segmentation tasks. Liu et al. (2021) presented a few short-learning model for intracranial hemorrhage segmentation model with classification task as the auxiliary task. These studies show that segmentation tasks can help highlight the regions with hemorrhage and extract more discriminative features (Xie et al., 2020). However, the segmentation task relies on highly accurate pixel-level annotations, which are time-consuming and challenging for large scale datasets. Furthermore, some researchers directly applied the object detection models for the diagnosis of hemorrhage. Chang et al. (2018) collected 11,021 CT scans from a single institution and proposed a hybrid convolutional neural network (CNN) with mask RCNN (He et al., 2017) for ICH detection and quantification. However, the regions with hemorrhage are fewer in the CT scans, and most CT scans have no hemorrhage. These anchor-based detection methods face the imbalance problem of samples and are always difficult for model optimization.

In fact, the region with hemorrhage is the most important basis for distinguishing blend sign from normal ICH. Radiologists mainly make evaluations based on the region with hemorrhage (Li et al., 2017). Then we employ the region with hemorrhage as a prior knowledge, and hope to add this prior knowledge to the model. This study proposes a prior knowledge-based model for the precise diagnosis of blend sign from head CT scans. We apply object detection to replace segmentation as the auxiliary task to reduce the annotation difficulty. Object detection only needs region-level annotation, e.g., center and bounding box, which are simpler and more robust than segmentation. Furthermore, there are inevitably some inaccuracies in the annotations. We propose a self-knowledge distillation strategy to deal with the inaccuracy annotations. We train a model as the teacher model and generate the pseudo labels for the training images. The pseudo labels contain much information about the negative categories. Using the pseudo labels as the supervision can gradually reduce the impact of inaccuracy annotations. Finally, we evaluated the proposed model on the collected dataset. Extensive results show that our method achieves better performance than the baseline model.

The contributions of this study are 3-fold:

- We retrospectively collected 1,749 anonymous non-contrast head CT scans from the First Affiliated Hospital of China Medical University and annotated 13,276 slices for evaluation. These slices are divided into three categories: no intracranial hemorrhage (non-ICH), normal intracranial hemorrhage (normal ICH), and blend sign.
- We present a prior knowledge-based diagnosis of blend sign from head CT scans. We apply an object detection task as the auxiliary task to reduce the annotation difficulty. The auxiliary task can help the model pay more attention to the regions with hemorrhage.
- We propose a self-knowledge distillation strategy to deal with incorrect annotations. The soft predictions contain much information about the negative categories and can gradually reduce the impact of inaccuracy annotations.

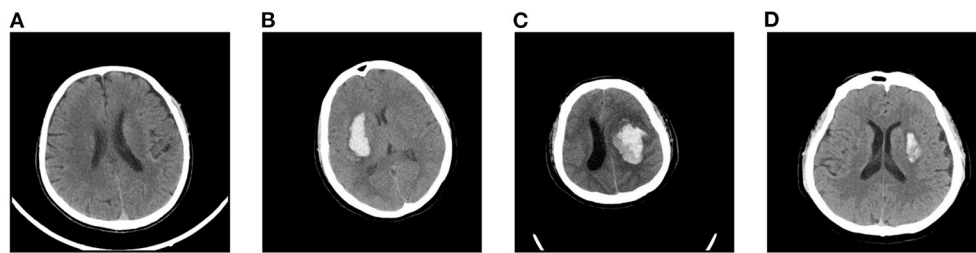


FIGURE 1
Examples of slices in our dataset: (A) non-ICH, (B) normal ICH, (C) blend sign, and (D) blend sign mimic (Normal ICH).

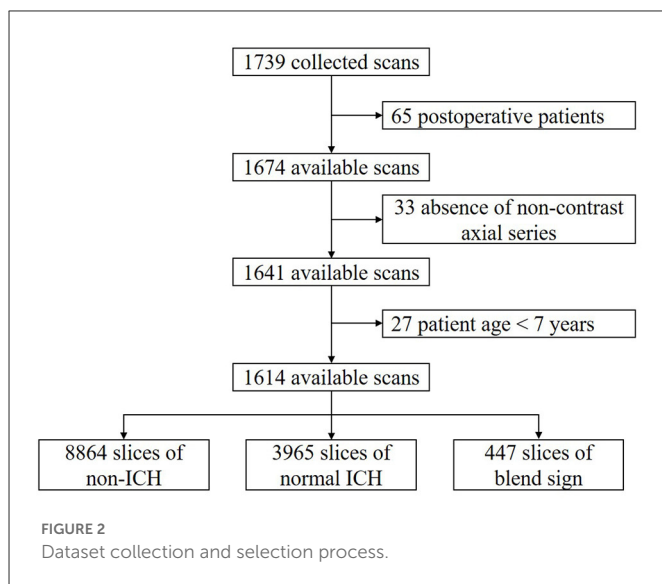


FIGURE 2
Dataset collection and selection process.

Materials and methods

Data collection and labeling

We retrospectively collected 1,749 anonymous non-contrast head CT scans from the Department of Neurosurgery, the First Hospital of China Medical University, Shenyang, China. The time of data collection is from 19 September 2018 to 24 December 2020. All scans were from the Asian population. CT scanners used in our dataset had slices per rotation varying from 16 to 128. All of the CT scans in our dataset were independently annotated at scan level and slice level by three radiologists. These radiologists had corresponding experiences of 6, 11, and 15 years in interpreting head CT scans. None of them was involved in the clinical care of the enrolled patients. After careful review and annotation, 125 CT scans were then excluded from further analysis due to the following reasons: postoperative patients (65); absence of non-contrast axial series (33); and patients were younger than 7 years (27). The remaining 1,614 available CT scans were finally used in our study. The dataset collection and selection process is shown in Figure 2.

Each of the experienced radiologists independently evaluated the scans and slices in our dataset. For the classification task, experienced radiologists recorded the following findings for each slice (Chilamkurthy et al., 2018): (1) normal head CT (non-ICH), (2) normal ICH (including EDH, SDH, SAH, IPH, and IVH), (3)

blend sign. In some slices, both normal ICH and blend sign may occur in the same slices, and we consider these slices as blend sign slices. For the detection task, each radiologist records whether one slice contains hemorrhage or not. For each hemorrhage region, the experienced radiologists annotate the center point and its bounding box (left, right, top, and bottom).

We apply the majority vote of these three radiologists' interpretations as the gold standard (Chilamkurthy et al., 2018). By slice-level annotation, there were 8,864 slices with non-ICH, 3,965 slices with normal ICH, and 447 slices with blend sign. The ratio of blend sign: normal ICH:non-ICH is approximately equal to 1 : 9 : 22. To effectively evaluate our algorithm as well as benefit the learning process, we intentionally kept such a high prevalence of blend sign in our dataset to ensure that there were enough positive samples.

Data pre-processing and augmentation

To highlight and emphasize specific pixels, we choose three different windows and encode them into the following RGB images: tissue window (WL = 40, WW = 40) for the red channel; brain window (WL = 50, WW = 100) for the green channel; and blood window (WL = 60, WW = 40) for the blue channel. Before being fed into the model, we reshape all CT slices to size 512×512 to reduce GPU memory usage. Then we convert all pixels from CT slices into floating point tensors and rescale the pixels values (between 0 and 255) into the [0, 1] interval.

Considering that our dataset is relatively small, we apply data augmentation to mitigate over-fitting in our task. In this article, we choose five forms of data augmentation operations, left-right flipping, left-right shifting, up-down shifting, random rotations (up to 10 degrees), and random scaling (0.9 to 1.1). The augmentation operations are shown in Figure 3. During the training process, the data generator will randomly choose the above augmentation operations for each slice, which means that the input to the model is different at each epoch. We find that data augmentation could largely enrich the training dataset and improves the performance of our model on the task of blend sign and normal ICH detection.

Overview of the proposed method

The proposed method consists of three parts: the pre-trained DCNN, the classification branch, and the detection branch. The pipeline of the proposed method is shown in Figure 4. The input

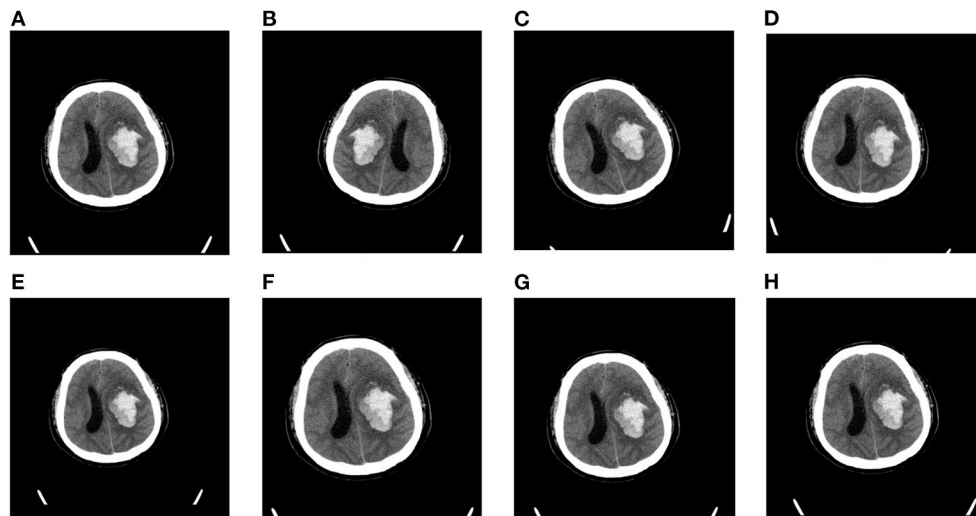


FIGURE 3
Random augmentations for CT slices: (A) original slice, (B) left-right flipping, (C) rotation (10°), (D) rotation (−10°), (E) scaling (0.9), (F) scaling (1.1), (G) shifting (right), and (H) shifting (down).

image is fed into the pre-trained DCNN to extract the feature maps. Then the feature maps from the last convolutional block (Block L) are fed into two branches, the classification branch and the detection branch. The classification branch gets the predictions with global average pooling (GAP) layer and linear layers. Then detection branch gets the locations with convolutional layers. In this way, the detection branch could transfer the hemorrhage localization information to boost the classification branch.

Pre-trained networks

Considering that our dataset is relatively small, we decided to use some pre-trained deep convolutional neural networks (DCNNs) as the backbones. Following (Lee et al., 2019), we choose one widely applied architecture, VGG16 (Simonyan and Zisserman, 2015), ResNet-50 (He et al., 2016), and Inception v3 (Szegedy et al., 2016), as the pre-trained networks. These architectures are pre-trained on one subset of ImageNet dataset (Deng et al., 2009). For the classification branch, we add one GAP layer and one linear layer after the last convolutional block (block L). We add two convolutional layers for the detection branch after the last convolutional block (Block L). Then we fine-tune the pre-trained backbone, the classification branch, and the detection branch with our dataset.

Classification branch

Automatic bleed sign classification is a multi-class classification problem. Each input image can be labeled as three mutually exclusive types. For each input slice, the classification branch will produce a three-dimensional output vector, where output_i is the probability that the input slice belongs to class i . The conventional loss is cross-entropy (CE) loss as follows:

$$L_{ce} = - \sum_{i=1}^3 y^i \ln \hat{y}^i \quad (1)$$

Where y^i is the ground truth and \hat{y}^i is the probability of class i our model predicts given an input slice x .

Considering the class-imbalanced problem in our dataset, models trained on these samples are biased toward dominate classes, non-ICH and normal ICH. To deal with the class-imbalanced problem, we try to use weighted cross-entropy (WCE) loss function as:

$$L_{wce} = - \sum_{i=1}^3 w_i y^i \ln \hat{y}^i \quad (2)$$

Where w_i is the corresponding loss weight for class i . The weights are to reduce the effects of imbalanced data distribution. The weights can be fixed or automatically adjusted during the training process. Through conductive experiment, we set the weights as $w_{1,2,3} = 1, 2, 20$, respectively. In the experiments, we find that WCE loss could partially solve the class-imbalanced problem and improve the performance of our algorithm.

Detection branch

The detection branch is inspired by the famous anchor-free detection frameworks, such as FCOS (Tian et al., 2019) and CenterNet (Duan et al., 2019). The detection branch aims to predict the localization of the hemorrhage region for the input image. In this branch, we consider the normal ICH and bleed sign as the foreground, and consider the non-ICH as the background. The output of the detection branch is one of the key-point maps with five channels. The first channel indicates the probability of the center point of the hemorrhage region. The other four channels are the bounding box (left, right, top, and bottom) of the hemorrhage region. The loss function for the detection branch is defined as:

$$L_{det} = \frac{1}{N} \sum_{j=1}^N L_{cls}(p_j, \hat{p}_j) + \frac{1}{N} \sum_{j=1}^N p_j L_{reg}(b_j, b_j^*), \quad (3)$$

$$L_{cls}(p_j, \hat{p}_j) = p_j \ln \hat{p}_j + (1 - p_j) \ln \hat{p}_j, \quad (4)$$

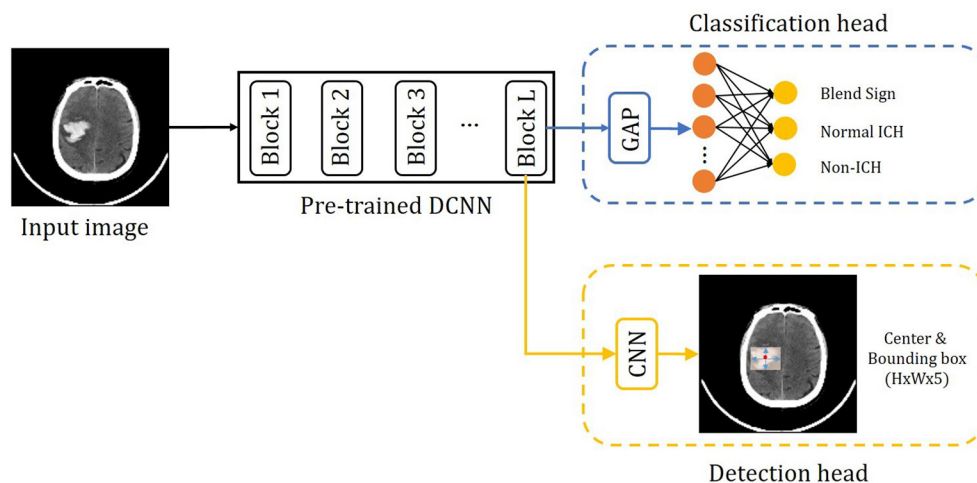


FIGURE 4

Overview of the precise diagnosis of the blend sign system. Given one CT slice, we first use a pre-trained CNN (trained on the ImageNet dataset) to extract feature maps, then put the feature maps into the classification branch and the segmentation branch, respectively.

$$L_{reg}(b_j, \hat{b}_j) = |b_j - \hat{b}_j|, \quad (5)$$

Where p_j and \hat{p}_j are the ground truth and probability of the j -th position being the center point of one hemorrhage region. The ground truth p_j is equal to 1 if the j -th position is the center point of one hemorrhage region, and p_j is equal to 0 if the j -th position is not the center point of one hemorrhage region. b_j represents the ground truth bounding box ($b_{j,left}$, $b_{j,right}$, $b_{j,top}$, and $b_{j,bottom}$) for the hemorrhage region, \hat{b}_j is the predicted bounding box. N is the number of pixels in an input image. L_{cls} is log loss over two classes (center point vs. not center point). L_{reg} is the smooth L_1 loss following (Tian et al., 2019). Following Duan et al. (2019), we perform Gaussian rendering for all ground truth p for faster training.

Self-knowledge distillation

For some slices, it is difficult to distinguish blend sign and normal ICH. Thus, it will inevitably lead to some inaccurate labels. The inaccurate annotations may bring a certain degree of challenges for the training process and the model would be misled by these inaccurate labels. We applied a self-knowledge distillation (Zhang et al., 2021) strategy to solve this problem. In addition to the positive category, the predictions also contain a lot of information about the negative category. The information on the negative category can be transferred to the student by self-knowledge distillation. Thus, the inaccuracy annotations can be changed gradually. Figure 5 shows the training process of self-knowledge distillation. We first train a model with the training images as the teacher model. Then we use the teacher model to generate predictions (pseudo labels) for each training image. Next, we train a student model by minimizing the distance between the predictions from the teacher and the student as follows:

$$L_{skd} = \text{KL}(\hat{y}_t || \hat{y}_s), \quad (6)$$

Where KL is the Kullback–Leibler divergence to measure the distance between two predictions. \hat{y}_t and \hat{y}_s are the predictions of the classification branch from the teacher and the student, respectively.

Since inaccurate labels mainly affect the classification branch, we apply self-knowledge distillation for the classification task. We apply the student model as the teacher model in the next iteration of the training process.

Optimization

The whole loss function for the proposed method is as follows:

$$L = L_{wce} + \lambda_1 L_{det} + \lambda_2 L_{skd}, \quad (7)$$

Where λ_1 and λ_2 are the hyper-parameters to balance the effect of different losses. Through conductive experiment, we set $\lambda_1 = 1$ and $\lambda_2 = 10$ for our experiments.

Experiments and results

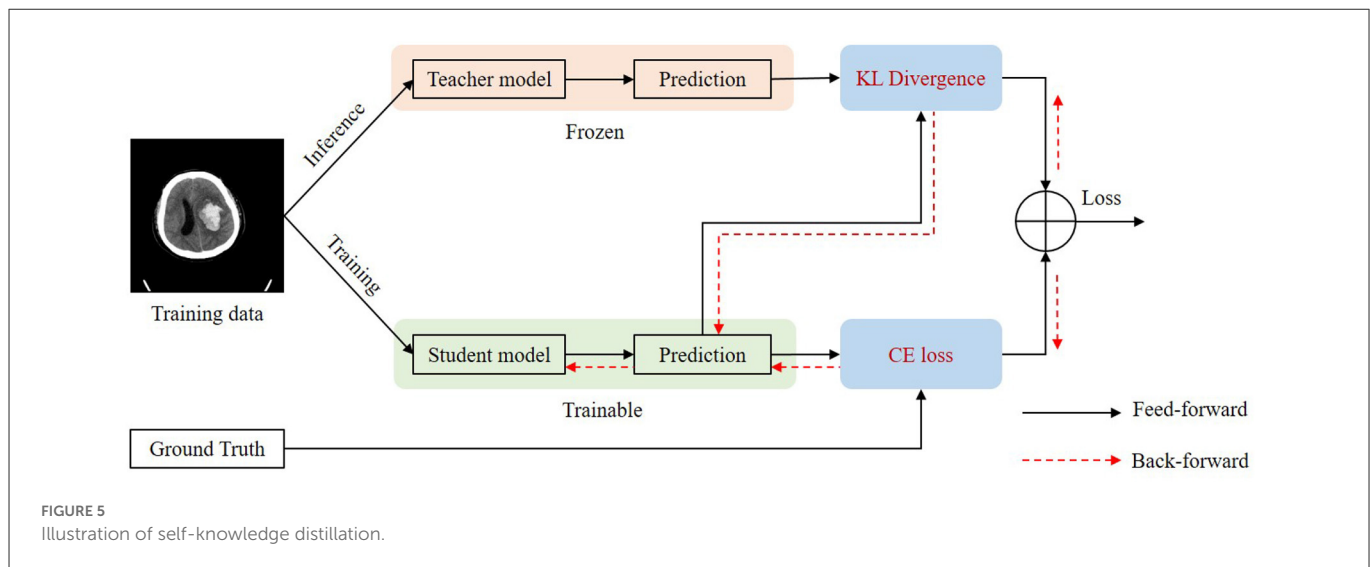
Implementation details

Training details

We carry out all experiments on the PyTorch framework (version 1.7.1) with a single Nvidia GTX 1080Ti GPU card of 11 GB memory. We choose stochastic gradient descent (SGD) with the momentum (0.9) and the weight decay (0.0005) as the optimizer. We apply the “poly” learning rate decay, in which the learning rate equals to $base_lr * (1 - \frac{iter}{total_iter})^{power}$. We set the base learning rate ($base_lr$) as 0.01 and the $power$ as 0.9. The mini-batch size is 16, the input size is 512×512 , and the training epoch number is 50. Data augmentation operations include random flipping, re-scaling (from 2 to 0.5), and rotation.

Evaluation metric

Considering that our dataset is with a class-imbalanced problem, we would like to choose *Sensitivity*, *Specificity*, *F1* score, and AUC (area under the receiver operating characteristic (ROC) curve) as the statistical evaluation metrics. We define TP as the number of true



positives, FP as the number of false positives, TN as the number of true negatives, and FN as the number of false negatives. The definition of sensitivity, specificity, and F1 score are as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad (8)$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (9)$$

$$F1 = \frac{2 \times TP}{2 \times TP + FN + FP}. \quad (10)$$

ROC curves were obtained by varying the threshold and plotting the true positive rate (i.e., sensitivity) and false positive rate (i.e., 1-specificity) at each threshold. We performed all statistical analyzes with the python package scikit-learn, and generated all statistical plots with Matplotlib.

Evaluation protocol

Considering that our dataset is relatively small, we may have insufficient samples for validation and test set if we use simple hold-out validation. The evaluation results may also be not reliable. To address this problem, we apply K -fold cross-validation as the evaluation protocol. It will split the dataset into K partitions with equal size. For each fold, we choose one partition as the validation set and train a model on the remaining $K-1$ partitions. The final score would then be the average of K validation scores. Particularly, in this study, we use a 5-fold validation as the evaluation protocol. We will randomly split the dataset into five partitions of equal size, and then train and evaluate five different models. Then the final evaluation score is the average of five different evaluation scores.

Results

Table 1 presents the comprehensive comparisons with existing datasets on five aspects: number of scans, five-type annotation, blend

TABLE 1 Comparison with other related datasets for the detection of intracranial hemorrhage.

Dataset	# Scans	Five-type annotation	Pixel-wise annotation	Blend sign
Chilamkurthy et al. (2018)	491	✓	×	×
Grewal et al. (2018)	252	×	✓	×
Lee et al. (2019)	904	✓	×	×
Ye et al. (2019)	2,836	✓	×	×
Ours	1,614	×	×	✓

*Scans: number of scans in the dataset. Five-type annotation: annotation for ICH, IPH, IVH, EDH, and SDH.

sign, and pixel-wise annotation. We can find that most datasets focus on the five-type (ICH, IPH, IVH, EDH, and SDH) detection tasks. These datasets rarely provide the pixel-wise annotation except for the dataset, as explained by in Grewal et al. (2018). This is because the pixel-wise annotation is time-consuming, which is challenging for large scale datasets. Our dataset is the only one to consider the detection of blend sign.

Table 2 presents the inter-rater interpretation agreement among the three radiologists. Concordance between the three radiologists on our dataset was higher for non-ICH (All Fleiss's $\kappa = 0.91$), representing excellent agreement with these findings. Blend sign has the lowest concordance with All Fleiss's $\kappa = 0.79$, indicating substantial agreement.

Table 3 summarizes the performance of the proposed method. Our method achieved AUCs of 0.972 for blend sign, 0.978 for normal ICH, and 0.999 for non-ICH. For Sensitivity, our method achieved 0.845 for blend sign, 0.898 for normal ICH, and 0.984 for non-ICH. For Specificity, our method achieved 0.941 for

TABLE 2 Reliability of the gold standards for our dataset.

	R1 and R2		R2 and R3		R1 and R3		All Fleiss's κ
	$p(\%)$	κ	$p(\%)$	κ	$p(\%)$	κ	
Non-ICH	97	0.91	98	0.92	98	0.91	0.91
Normal ICH	94	0.85	96	0.87	95	0.86	0.86
Blend sign	89	0.78	92	0.81	91	0.79	0.79

R, radiologist; p , percentage agreement rate; κ , Cohen's κ coefficient, a statistic to measure inter-rater agreement; All Fleiss's κ coefficient, a statistic to measure the agreements among multiple raters. Values of κ and All Fleiss's κ > 0.80 indicate excellent agreement, and 0.60–0.80 indicate substantial agreement.

TABLE 3 Performance of the automated detection algorithm on our dataset with ResNet-50 as the backbone.

	WCE	AT	SKD	Sensitivity	Specificity	F1 score	AUC
	✓			0.544	0.986	0.672	0.966
Blend Sign	✓	✓		0.716	0.977	0.768	0.972
	✓	✓	✓	0.845	0.941	0.781	0.977
	✓			0.929	0.842	0.913	0.964
Normal ICH	✓	✓		0.897	0.897	0.911	0.967
	✓	✓	✓	0.898	0.936	0.928	0.978
	✓			0.978	0.953	0.932	0.996
Non-ICH	✓	✓		0.982	0.945	0.929	0.994
	✓	✓	✓	0.984	0.986	0.968	0.999

WCE, weighted cross-entropy loss; AT, auxiliary task (detection task); SKD, self-knowledge distillation. The bold values indicate the best score.

blend sign, 0.936 for normal ICH, and 0.986 for non-ICH. Based on these results, we have three significant findings. First, the performance of normal ICH was consistently higher than that of the blend sign. This may be because, compared with the blend sign, slices with normal ICH are much easier to be discriminated. Second, auxiliary task (detection task) and self-knowledge distillation method could boost the performance, especially for the sensitivity of blend sign (from 0.544 to 0.845). This demonstrated the effectiveness of auxiliary task (detection task) and self-knowledge distillation method. Third, the sensitivity of normal ICH was slightly decreased from 0.929 to 0.898. This is because the model would predict some slices with normal ICH into blend sign. Table 4 shows the performance with different pre-trained backbones. We can observe that there are slight differences among different pre-trained backbones.

TABLE 4 Performance of the automated detection algorithm with different pre-trained backbones.

	Backbone	F1 score	AUC
	VGG16	0.762	0.964
Blend sign	ResNet-50	0.781	0.977
	Inception V3	0.773	0.970
	VGG16	0.932	0.981
Normal ICH	ResNet-50	0.928	0.978
	Inception V3	0.915	0.972
	VGG16	0.949	0.995
Non-ICH	ResNet-50	0.968	0.999
	Inception V3	0.955	0.997

Error analysis

We introduce the confusion matrix to evaluate the performance and analyze the error. The row of the confusion matrix is the predicted label, and the column is the true label. Figures 6A–C shows the confusion matrix with different combinations. From Figure 6A, we find that the classification of blend sign is terrible, 43% of slices with blend signs were incorrectly predicted as non-ICH. Meanwhile, only 2% of slices with normal ICH were incorrectly predicted as blend sign. It is because the model is unable to distinguish the slices with normal ICH and blend sign. The model tends to predict the slice with a blend sign as normal ICH. In addition, we propose

the auxiliary task (detection task) and find that only 28% of slices with blend sign were incorrectly predicted as non-ICH. The results indicate that the auxiliary task (detection task) can force the model to pay more attention to the regions with hemorrhage and make more accurate predictions. Moreover, when we apply the self-knowledge distillation, 85% of slices with blend sign were correctly predicted, and only 15% of slices were incorrectly predicted as non-ICH. The improvement in prediction accuracy demonstrates the effectiveness of the self-knowledge distillation method for inaccurate annotations. The above-mentioned results indicate that the main challenge is to distinguish blend sign from normal ICH. Our proposed auxiliary task

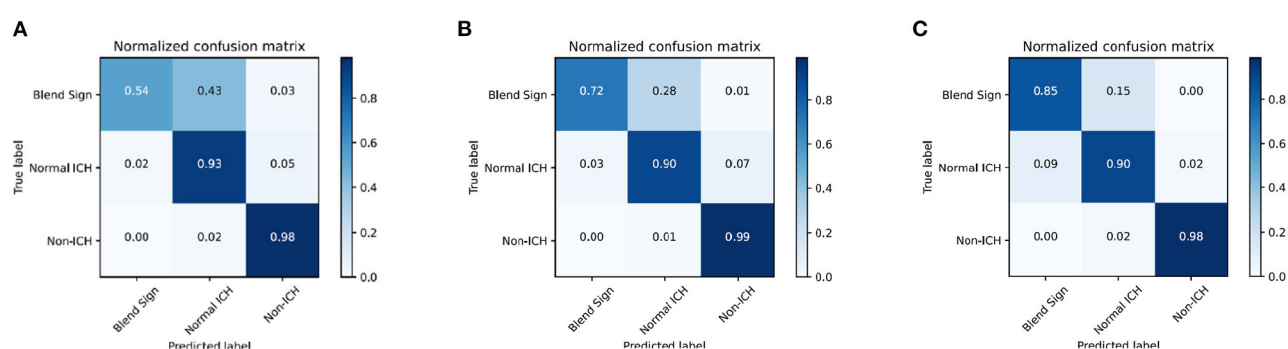


FIGURE 6 Confusion matrix of the classification results. **(A)** WCE, **(B)** WCE+AT, and **(C)** WCE+AT+SKD. WCE, weighted cross-entropy loss; AT, auxiliary task (detection task); SKD, self-knowledge distillation.

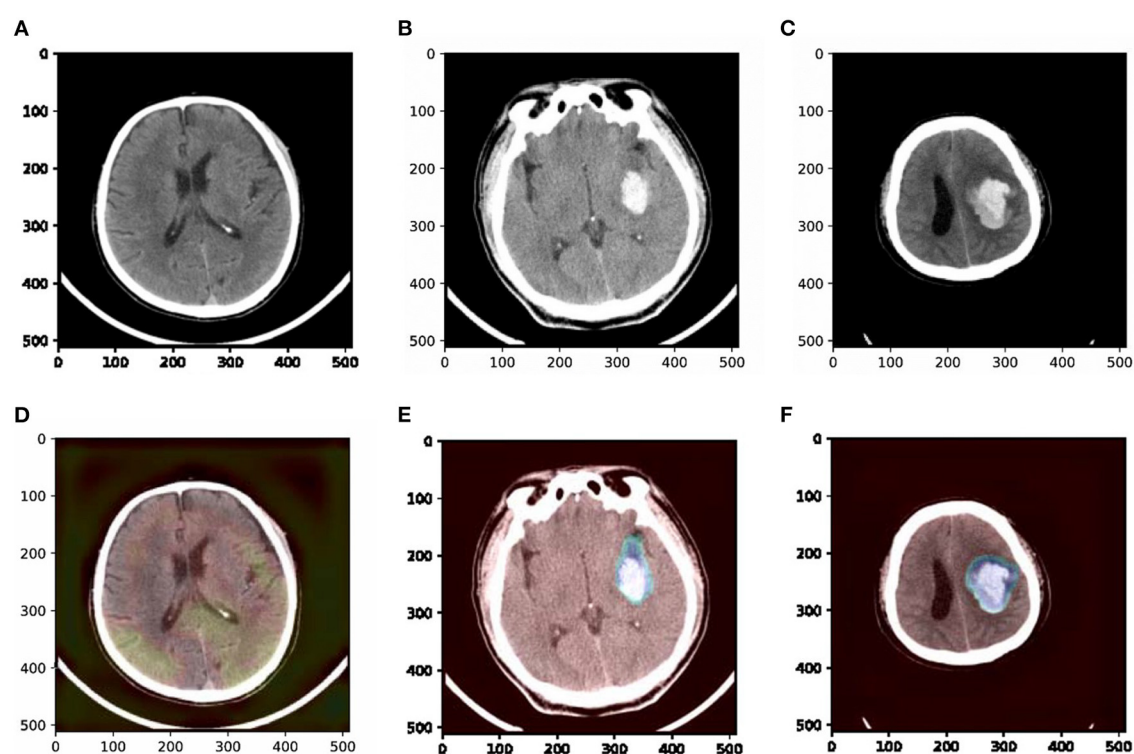


FIGURE 7 Examples of original slices with their corresponding class activation heatmaps. **(A)** Original slice of non-ICH, **(B)** original slice of normal ICH, **(C)** original slice of blend sign, **(D)** class activation heatmaps of the original slice of non-ICH, **(E)** class activation heatmaps of the original slice of normal ICH, and **(F)** class activation heatmaps of the original slice of blend sign. It is interesting to note that the regions with hemorrhage are strongly activated, and this is probably how the network can make decisions.

(detection task) and self-knowledge distillation can partially solve this challenge.

Visualizing what our model learns

To understand which parts of a given image lead to the final classification decision, we apply “Class Activation Map” (CAM) visualization as a powerful technique (Ye et al., 2019). The specific method we choose is Grad-CAM (Selvaraju et al., 2017). CAM

visualization produces heatmaps of “class activation” over input images. These heatmaps are 2D score grids, which indicate how important each location is regarding the class considered. CAM visualization is helpful to understand the decision process of our model as well as analyze the classification errors. It also might provide guidance for interpretation during clinical applications. Three examples from our dataset are shown in Figure 7, where the bright areas indicated high importance for decision marking and gray areas indicated low importance. It was interesting to note that the areas with bleeding attracted the most attention, while the areas

without hemorrhage attracted less attention. In addition, we can also observe some non-overlapping regions between the highlighted regions and the bleeding regions. These results demonstrate that our approach could partly guide the model to pay more attention to the bleeding regions and misclassify some bleeding regions. The constraints of the detection task may cause misclassification. The detection task could only encourage the model to pay more attention to some rectangular areas rather than the accurate pixels.

Discussion

This study focuses on the precise diagnosis of blend sign from head CT scans with a deep learning approach. Our contributions are summarized in the following aspects. First, we retrospectively collected 1,749 anonymous non-contrast head CT scans and annotated 13,276 slices for evaluation. Second, we present a prior knowledge-based model for blend sign from head CT scans. We apply an object detection task as the auxiliary task to help the model pay more attention to the regions with hemorrhage. Third, we propose a self-knowledge distillation strategy to reduce the impact of incorrect annotations. Results from Figure 6 and Table 3 confirm that the proposed auxiliary task (object detection task) and self-knowledge distillation indeed improved the performance of blend sign detection. Coarse heatmaps in Figure 7 show that the regions with hemorrhage attract more attention than the other regions. These heatmaps have the potential to be employed as a coarse bleeding localization map. In summary, our proposed algorithm assists the detection of blend sign and normal ICH with high accuracy, which may be a useful tool for the precise diagnosis of blend sign.

The proposed algorithm produces a pretty good performance in our dataset. AUCs for all the findings were >0.97 . The F1 score for all the findings except the blend sign is >0.92 . The *Specificity* for all the findings is greater than or equal to 0.94. For the diagnosis of blend sign, our algorithm achieves *Sensitivity* as 0.845, *Specificity* as 0.941, F1 score as 0.781, and AUC as 0.972. This may be because of two reasons. First, we retrospectively collected a large amount of anonymous non-contrast head CT scans from the First Affiliated Hospital of China Medical University. We apply the majority vote of three radiologists as the ground truth, which is a better gold standard than one radiologist. The quality of annotation is relatively high. Second, we apply an auxiliary task (object detection) and the self-knowledge distillation strategy, which are suitable for our condition. The auxiliary task could help the model to pay more attention to the regions with hemorrhage and extract more discriminate features. At the same time, self-knowledge distillation could vastly reduce the impact of incorrect annotations.

Our study also has several limitations. First, all slices in our dataset were from the Asian population, which may limit the generalization of our algorithm. It is desirable to include information on populations from other continents in the future. Second, to enhance our algorithm's performance and ensure there are enough positive samples to train the model. The prevalence of ICH (including blend sign and normal ICH) in our dataset is much higher than in some popular datasets and real clinical diagnoses. For example, the reported incidence rate of ICH is 12% in the famous Quer25k dataset (Chilamkurthy et al., 2018), while in our dataset, the incidence rate of ICH is 32%. The performance of our algorithm may change in real clinical applications. Third, we only have 1,614 available cases

in our dataset, including 8,864 slices of non-ICH, 3,965 slices of normal ICH, and 447 slices of blend sign. The number of blend sign is quite limited. Performance may be adversely affected by the lack of training examples. Although transfer learning, auxiliary task, and self-knowledge distillation could boost the performance of our algorithm, the performance may drop a lot in real clinical cases. The next step is expanding our dataset and collecting more available scans, especially with blend sign. Finally, Fleiss's κ coefficient for blend sign is just 0.79, which means there is some inconsistency in the annotation of many slices with blend sign. The low blend sign identification rate of junior radiologists may need more investigation and may affect the training and generalization of our algorithm. Although self-knowledge distillation could partially alleviate the impact of inaccurate labeling, we also need to improve the reliability of annotations in the future.

Conclusion

In this study, we propose a prior knowledge-based precise diagnosis of blend sign from head CT scans. We constructed a dataset with 1,614 available cases, 8,864 slices with non-ICH, 3,965 slices with normal ICH, and 447 slices with blend sign. To better distinguish the slices with normal ICH and blend sign, we propose the object detection task as an auxiliary task in addition to the classification task. The auxiliary task can help the model pay more attention to the region with hemorrhage. In addition, we employ a self-knowledge distillation strategy to reduce the influence of inaccurate annotations. Our precise diagnosis may assist less-experienced head CT interpreters in reducing initial misinterpretations. It also may reduce radiologists' workload and improve efficiency in a natural clinical setting. Experimental results show that our method achieved AUCs of 0.972 for blend sign, 0.978 for normal ICH, and 0.999 for non-ICH, which is a pretty good performance. In the future, we plan to collect more available scans with high reliability of annotations and extend our algorithm to clinical practice.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

CW contribute to conceptualization, methodology, software, and original draft writing. JY contributes to data collection and annotation. JZ contributes to conceptualization, supervision, resources, and funding acquisition. SH contributes to conceptualization and data annotation. YQ contributes to software and validation. BF contributes to conceptualization and review. XL

contributes to investigation and review. All authors contributed to the article and approved the submitted version.

Funding

This work was partially supported by the National Natural Science Foundation of China (62176029 and 61876026), the National Key Research and Development Program of China (2017YFB1402401), and the Key Research Program of Chongqing Science and Technology Bureau (cstc2020jscx-msxmX0149).

Acknowledgments

The authors thank Yupin Yang and Ganghong Huang for their great helps with codes, suggestions, and discussion.

References

- Alfaro, D., Levitt, M. A., English, D. K., Williams, V., and Eisenberg, R. (1995). Accuracy of interpretation of cranial computed tomography scans in an emergency medicine residency program. *Ann. Emerg. Med.* 25, 169–174. doi: 10.1016/S0196-0644(95)70319-5
- Chang, P. D., Kuoy, E., Grinband, J., Weinberg, B. D., Thompson, M., Homo, R., et al. (2018). Hybrid 3d/2d convolutional neural network for hemorrhage evaluation on head ct. *Am. J. Neuroradiol.* 39, 1609–1616. doi: 10.3174/ajnr.A5742
- Chen, X., Lu, Y., Bai, J., Yin, Y., Cao, K., Li, Y., et al. (2018). “Train a 3D U-Net to segment cranial vasculature in cta volume without manual annotation,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC: IEEE), 559–563.
- Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., et al. (2018). Deep learning algorithms for detection of critical findings in head ct scans: a retrospective study. *Lancet* 392, 2388–2396. doi: 10.1016/S0140-6736(18)31645-3
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). “ImageNet: a large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL: IEEE), 248–255.
- Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., and Tian, Q. (2019). “CenterNet: keypoint triplets for object detection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Seoul: IEEE).
- Grewal, M., Srivastava, M. M., Kumar, P., and Varadarajan, S. (2018). “RadNet: radiologist level accuracy using deep learning for hemorrhage detection in CT scans,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC: IEEE), 281–284.
- Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., et al. (2017). Brain tumor segmentation with deep neural networks. *Med. Image Anal.* 35, 185–197. doi: 10.1016/j.media.2016.05.004
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). “Mask R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision* (Venice: IEEE), 2961–2969.
- 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* (Las Vegas, NV: IEEE), 770–778.
- Heit, J. J., Iv, M., and Wintermark, M. (2017). Imaging of intracranial hemorrhage. *J. Stroke* 19, 11. doi: 10.5853/jos.2016.00563
- Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., and Shen, Y. (2021). Bidirectional mapping generative adversarial networks for brain mr to pet synthesis. *IEEE Trans. Med. Imaging* 41, 145–157. doi: 10.1109/TMI.2021.3107013
- Hu, S., Yuan, J., and Wang, S. (2019). “Cross-modality synthesis from MRI to pet using adversarial U-Net with different normalization,” in *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)* (Shenzhen: IEEE), 1–5.
- Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2017). Efficient multi-scale 3D CNN with fully connected crf for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78. doi: 10.1016/j.media.2016.10.004
- Kuo, W., Hane, C., Mukherjee, P., Malik, J., and Yuh, E. L. (2019). Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proc. Natl. Acad. Sci. U.S.A.* 116, 22737–22745. doi: 10.1073/pnas.1908021116
- Lee, H., Yune, S., Mansouri, M., Kim, M., Tajmir, S. H., Guerrier, C. E., et al. (2019). An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat. Biomed. Eng.* 3, 173–182. doi: 10.1038/s41551-018-0324-9
- Li, Q., Dong, F., Wang, Q., Xu, F., and Zhang, M. (2021). A model comprising the blend sign and black hole sign shows good performance for predicting early intracerebral haemorrhage expansion: a comprehensive evaluation of ct features. *Eur. Radiol.* 31, 9131–9138. doi: 10.1007/s00330-021-08061-y
- Li, Q., Yang, W.-S., Wang, X.-C., Cao, D., Zhu, D., Lv, F.-J., et al. (2017). Blend sign predicts poor outcome in patients with intracerebral hemorrhage. *PLoS ONE* 12, e0183082. doi: 10.1371/journal.pone.0183082
- Li, Q., Zhang, G., Huang, Y.-J., Dong, M.-X., Lv, F.-J., Wei, X., et al. (2015). Blend sign on computed tomography: novel and reliable predictor for early hematoma growth in patients with intracerebral hemorrhage. *Stroke* 46, 2119–2123. doi: 10.1161/STROKEAHA.115.009185
- Li, Q., Zhang, G., Xiong, X., Wang, X.-C., Yang, W.-S., Li, K.-W., et al. (2016). Black hole sign: novel imaging marker that predicts hematoma growth in patients with intracerebral hemorrhage. *Stroke* 47, 1777–1781. doi: 10.1161/STROKEAHA.116.013186
- Li, Y.-H., Zhang, L., Hu, Q.-M., Li, H.-W., Jia, F.-C., and Wu, J.-H. (2012). Automatic subarachnoid space segmentation and hemorrhage detection in clinical head ct scans. *Int. J. Comput. Assist. Radiol. Surg.* 7, 507–516. doi: 10.1007/s11548-011-0664-3
- Liu, X., Zhu, S., Yang, W., and Zhou, S. (2021). “Few-shot learning for automatic intracranial hematoma segmentation,” in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Houston, TX: IEEE), 3400–3407.
- Nguyen, H. S., Li, L., Patel, M., and Mueller, W. (2016). Density measurements with computed tomography in patients with extra-axial hematoma can quantitatively estimate a degree of brain compression. *Neuroradiol. J.* 29, 372–376. doi: 10.1177/1971400916658795
- Prevedello, L. M., Erdal, B. S., Ryu, J. L., Little, K. J., Demirer, M., Qian, S., et al. (2017). Automated critical test findings identification and online notification system using artificial intelligence in imaging. *Radiology* 285, 923–931. doi: 10.1148/radiol.2017162664
- Qureshi, A. I., Mendelow, A. D., and Hanley, D. F. (2009). Intracerebral haemorrhage. *Lancet* 373, 1632–1644. doi: 10.1016/S0140-6736(09)60371-8
- Qureshi, A. I., Tuhim, S., Broderick, J. P., Batjer, H. H., Hondo, H., and Hanley, D. F. (2001). Spontaneous intracerebral hemorrhage. *N. Engl. J. Med.* 344, 1450–1460. doi: 10.1056/NEJM200105103441907
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). “Grad-CAM: visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE International Conference on Computer Vision* (Venice: IEEE), 618–626.
- Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* 35, 1285–1298. doi: 10.1109/TMI.2016.2528162
- Simonyan, K., and Zisserman, A. (2015). “Very deep convolutional networks for large-scale image recognition,” in *Proceedings of International Conference on Learning Representations*.

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.

- Strub, W., Leach, J., Tomsick, T., and Vagal, A. (2007). Overnight preliminary head ct interpretations provided by residents: locations of misidentified intracranial hemorrhage. *Am. J. Neuroradiol.* 28, 1679–1682. doi: 10.3174/ajnr.A0653
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 2818–2826.
- Tian, Z., Shen, C., Chen, H., and He, T. (2019). “FCOS: fully convolutional one-stage object detection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Seoul: IEEE), 9627–9636.
- Titano, J. J., Badgeley, M., Schefflein, J., Pain, M., Su, A., Cai, M., et al. (2018). Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat. Med.* 24, 1337–1341. doi: 10.1038/s41591-018-0147-y
- Wang, S., Chen, Z., You, S., Wang, B., Shen, Y., and Lei, B. (2022). Brain stroke lesion segmentation using consistent perception generative adversarial network. *Neural Comput. Appl.* 34, 8657–8669. doi: 10.1007/s00521-021-06816-8
- Wang, S., Shen, Y., Zeng, D., and Hu, Y. (2018). “Bone age assessment using convolutional neural networks,” in *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)* (Chengdu: IEEE), 175–178.
- Xiao, F., Liao, C.-C., Huang, K.-C., Chiang, I.-J., and Wong, J.-M. (2010). Automated assessment of midline shift in head injury patients. *Clin. Neurol Neurosurg.* 112, 785–790. doi: 10.1016/j.clineuro.2010.06.020
- Xie, Y., Zhang, J., Xia, Y., and Shen, C. (2020). A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Trans. Med. Imaging* 39, 2482–2493. doi: 10.1109/TMI.2020.2972964
- Yagi, K., Tao, Y., Hara, K., Hirai, S., Takai, H., Kinoshita, K., et al. (2019). Does noncontrast computed tomography scan predict rebleeding after endoscopic surgery for spontaneous intracerebral hemorrhage? *World Neurosurg.* 127:e965–e971. doi: 10.1016/j.wneu.2019.04.010
- Yang, Y., Chen, S., Tan, D., Yao, R., Zhu, S., Jia, Y., et al. (2021). “Fusion branch network with class learning difficulty loss function for recognition of haematoma expansion signs in intracerebral haemorrhage,” in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Houston, TX: IEEE), 3448–3455.
- Ye, H., Gao, F., Yin, Y., Guo, D., Zhao, P., Lu, Y., et al. (2019). Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur. Radiol.* 29, 6191–6201. doi: 10.1007/s00330-019-06163-2
- You, S., Lei, B., Wang, S., Chui, C. K., Cheung, A. C., Liu, Y., et al. (2022). Fine perceptive gans for brain mr image super-resolution in wavelet domain. *IEEE Trans. Neural Netw. Learn. Syst.* 2022, 3153088. doi: 10.1109/TNNLS.2022.3153088
- Yu, W., Lei, B., Ng, M. K., Cheung, A. C., Shen, Y., and Wang, S. (2021). Tensorizing gan with high-order pooling for Alzheimer’s disease assessment. *IEEE Trans. Neural Netw. Learn. Syst.* 2021, 3063516. doi: 10.1109/TNNLS.2021.3063516
- Yu, W., Lei, B., Wang, S., Liu, Y., Feng, Z., Hu, Y., et al. (2022). Morphological feature visualization of Alzheimer’s disease via multidirectional perception gan. *IEEE Trans. Neural Netw. Learn. Syst.* 2022, 318369. doi: 10.1109/TNNLS.2021.3118369
- Zhang, L., Bao, C., and Ma, K. (2021). Self-distillation: towards efficient and compact neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 4388–4403. doi: 10.1109/TPAMI.2021.3067100
- Zhang, M., Chen, J., Zhan, C., Liu, J., Chen, Q., Xia, T., et al. (2020). Blend sign is a strong predictor of the extent of early hematoma expansion in spontaneous intracerebral hemorrhage. *Front. Neurol.* 11, 334. doi: 10.3389/fneur.2020.00334



OPEN ACCESS

EDITED BY

Fei He,
Coventry University, United Kingdom

REVIEWED BY

Sreedhar Kollem,
SR University, India
Hossam El-Din Moustafa,
Mansoura University, Egypt

*CORRESPONDENCE

Wushouer Silamu
✉ lizongren@stu.xju.edu.cn

RECEIVED 24 March 2023

ACCEPTED 20 April 2023

PUBLISHED 12 May 2023

CITATION

Zongren L, Silamu W, Shurui F and
Guanghui Y (2023) Focal cross transformer:
multi-view brain tumor segmentation model
based on cross window and focal
self-attention.
Front. Neurosci. 17:1192867.
doi: 10.3389/fnins.2023.1192867

COPYRIGHT

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

Focal cross transformer: multi-view brain tumor segmentation model based on cross window and focal self-attention

Li Zongren¹, Wushouer Silamu^{1*}, Feng Shurui² and Yan Guanghui²

¹School of Information Science and Engineering, Xinjiang University, Urumqi, China, ²School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China

Introduction: Recently, the Transformer model and its variants have been a great success in terms of computer vision, and have surpassed the performance of convolutional neural networks (CNN). The key to the success of Transformer vision is the acquisition of short-term and long-term visual dependencies through self-attention mechanisms; this technology can efficiently learn global and remote semantic information interactions. However, there are certain challenges associated with the use of Transformers. The computational cost of the global self-attention mechanism increases quadratically, thus hindering the application of Transformers for high-resolution images.

Methods: In view of this, this paper proposes a multi-view brain tumor segmentation model based on cross windows and focal self-attention which represents a novel mechanism to enlarge the receptive field by parallel cross windows and improve global dependence by using local fine-grained and global coarse-grained interactions. First, the receiving field is increased by parallelizing the self-attention of horizontal and vertical fringes in the cross window, thus achieving strong modeling capability while limiting the computational cost. Second, the focus on self-attention with regards to local fine-grained and global coarse-grained interactions enables the model to capture short-term and long-term visual dependencies in an efficient manner.

Results: Finally, the performance of the model on Brats2021 verification set is as follows: dice Similarity Score of 87.28, 87.35 and 93.28%; Hausdorff Distance (95%) of 4.58mm, 5.26mm, 3.78mm for the enhancing tumor, tumor core and whole tumor, respectively.

Discussion: In summary, the model proposed in this paper has achieved excellent performance while limiting the computational cost.

KEYWORDS

brain tumor segmentation, cross window, CNN, Transformer, focal self-attention

1. Introduction

Brain tumors represent new growths in the cranial cavity that are also known as intracranial tumors and brain cancer and originate from the brain, meninges, nerves, blood vessels and brain appendages, or from other tissues or organs via metastasis. Most of these growths can produce headache, intracranial hypertension, and focal symptoms. The incidence of brain tumors is 7–10 per 100,000 subjects, and more than half of such tumors are malignant. According to a study by the

World Health Organization (WHO), brain tumors have become one of the three major tumors endangering human health. The early identification and effective segmentation of brain tumors is very important if clinicians are to formulate treatment plans and improve the survival rates. However, at present, clinicians mainly segment brain tumors from nuclear magnetic resonance imaging (MRI) by hand; this practice is time consuming and also renders the accuracy of segmentation entirely dependent on the experience of the technician or physician. Therefore, convolutional neural networks (CNNs) (Long et al., 2015) and Transformer (Vaswani et al., 2017; Chen et al., 2021; Yuan et al., 2021) and other computer-aided diagnostic technologies are increasingly becoming a new trend with which to segment brain tumor images. Figure 1 shows that MRI data of different morphologies captured different pathological features of tumors.

The segmentation method is based on convolutional neural networks (CNNs) and has generated remarkable achievements in the field of medical image segmentation and other visual fields with its powerful characterization ability. However, CNNs are associated with limitations in global modeling or remote contextual interactions and spatial dependencies prevent further expansion of brain tumor segmentation, thus inspiring the use of Transformer and attention mechanism in medical imaging. Following the pioneering work of Transformer in the field of vision, Vision Transformer (Dosovitskiy et al., 2020) has created a general model in the field of natural language processing (NLP) and vision (Zheng et al., 2021). Several variants were subsequently developed, assisting the introduction of Transformer into medical image classification, target detection, medical image segmentation, and other fields. However, with the prosperity of Transformer in the visual field, many researchers found that although the full attention mechanism of Transformer played a significant role in global modeling or remote context interaction, it also generated problems associated with computational complexity secondary growth (Zhang et al., 2021). Moreover, due to high computational complexity and memory consumption, the full self-attention mechanism of Transformer cannot be applied to medical image segmentation.

To improve efficiency and reduce computational complexity, researchers have suggested replacing the full self-attention mechanism with a limited range of local window self-attention mechanisms. Furthermore, considering the information interaction between windows, shift operation is utilized (Liu et al., 2021, 2022; Cao et al., 2023) and information can be exchanged between nearby Windows, thus alleviating the problem of computational efficiency, at least to some extent. However, expansion of the receptive field in this way is

rather slow, and many windows need to be stacked to achieve global self-attention (Liang et al., 2021). For high-resolution image models, such as medical image segmentation, a large receptive field is particularly important as this can affect the local or remote acquisition of contextual information. In view of this, this paper proposes a multi-view brain tumor segmentation model based on cross window and focal self-attention which can retain computational complexity while achieving a large receptive field. Several innovations and major contributions were involved in the development of this new model.

- An innovative mechanism were used to extract characteristic input information from brain tumors, and rich local semantic information was extracted with fine-grained interactions. Then, global semantic information was captured with coarse-grained interactions. This effectively alleviated the problem of high computational complexity associated with the global self-attention mechanism.
- The characteristic information of brain tumor was extracted by cross window, and the self-attention weights within the window were learned from both horizontal and vertical directions by concurrent multiple self-attention mechanisms; then, their weights were concatenated. This expands the receptive field of self-attentional learning and balances the relationship between computational complexity and self-attentional learning ability in Transformer.
- Locally enhanced location coding was adopted to apply the location information to the linear projection value; then, the location information was directly merged into each Transformer block, effectively improving the accuracy of pixel level segmentation for brain tumors.
- The novelty model proposed was applied to the field of brain tumor segmentation and verified on Brats2019 and Brats2021 data sets. The experimental results showed that the model proposed in this paper has achieved excellent performance and outstanding clinical application value.

The sections of this paper are arranged as follows. In the second section, we introduce the existing literature related to this paper. In the third section, we elaborate the architecture of the focal cross window model. The fourth section provides verification of model by using two brain tumor data sets, while the final section summarizes the main contents of this paper and discusses future research and perspectives.

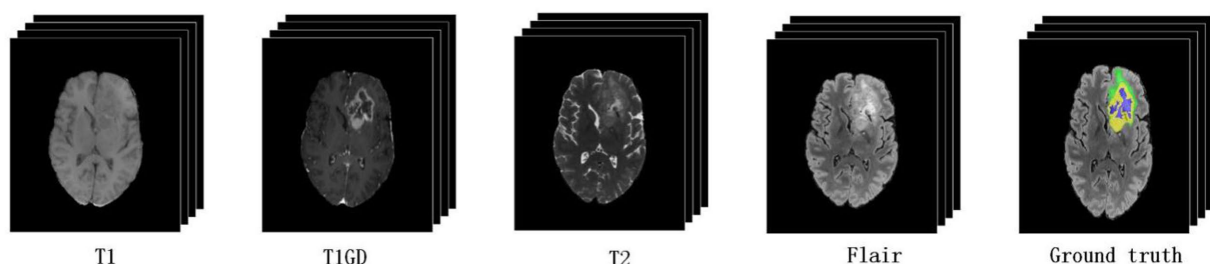


FIGURE 1

Magnetic resonance imaging (MRI) of multimodal brain tumors. The green, yellow, and blue regions in the ground truth indicate edema (ED), an enhancing tumor (ET), and non-enhancing tumor and necrosis (NCR/NET), respectively.

2. Related work

2.1. Vision Transformer

The Vision Transformer (Dosovitskiy et al., 2020) model, as the first application of Transformer in the field of computer vision, exhibits strong universality, not only in the field of NLP, but also in the field of vision. As far as possible, the model follows the design of the original Transformer model. Firstly, the two-dimensional input feature map was partitioned through the patch partition module, and the partitioned patch was flattened into a token sequence along the channel direction (Chu et al., 2021a,b). A learnable embedded token classification header was added to the original token sequence prior to self-attentional learning; this was implemented by a hidden layer perceptron (MLP) during pre-training (Chu et al., 2021a,b; Touvron et al., 2021; Zhu et al., 2021), implemented by a linear layer when fine-tuned. Because Transformer's self-attention learning sequence remains constant, it loses important location information. To solve this problem, researchers embedded the location coding information before multi-head self-attention learning. The model uses standard learnable 1D location embedding to preserve the location information in the token sequence. The encoder layer of Transformer is composed of multi-head attention and MLP modules, and the Layernorm (LN) layer is used before each module is applied (Gao et al., 2022; Huang et al., 2022; Lin et al., 2022). The groundbreaking results of the Vision Transformer model demonstrated that a pure Transformer-based architecture can achieve applications comparable to CNNs, thus demonstrating the potential of Vision Transformer for the unified processing of natural language processing and visual tasks. Influenced by the success of the Vision Transformer model, many researchers improved the model from a range of aspects, including computational complexity, segmentation accuracy, and parallelization, so as to improve the efficacy of downstream tasks such as target detection and image segmentation (Howard et al., 2017; He et al., 2021; Wang et al., 2021a,b; Yuan et al., 2021). This led to the development of the Swin Transformer model (Liu et al., 2021) which limits the self-attention learning scope of Vision Transformer to a local window and acquires global information by shifting information between local windows. Thus, the computational complexity of the model is reduced, and the accuracy of image classification is improved. Some researchers combined Vision Transformer with a CNN to connect input features with the Transformer layer after convolution processing, learn local information through CNN, learn global semantic information by Transformer, and then combine the two strategies. This allowed the acquisition of rich semantic feature information. However, when Swin Transformer switches information between local windows during shift operation, the receptive field expands slowly, and many Transformer blocks need to be stacked to obtain global semantic information. However, combining CNN with Transformer (Wang et al., 2021a,b) makes Transformer lose its original ability to capture short-term and remote semantic information at the same time. Therefore, to solve these above problems, we applied the Cross Window to balance the relationship between the computational complexity of the model and the self-attentional learning ability. Under the premise of reducing computational complexity, we expanded the receptive field of self-attentional learning, thus improving the accuracy of brain tumor segmentation.

2.2. The global and local self-attention

In the field of medical image analysis, Transformer models usually need to process many long sequence tokens due to the high resolution of images. Over recent years, many researchers have proposed various effective self-attention mechanisms to solve the problem of secondary computing and high memory overhead in Transformer. On the one hand, for many applications featuring medical image segmentation, CNN is combined with Transformer. The token quantity is reduced through CNN down-sampling, and then the global self-attention weight is acquired by coarse-grained interactions. Although this method can improve the efficiency of Transformer, it loses rich semantic information around the tokens, and loses the ability to capture both short-term and remote semantic information. On the other hand, fine-grained self-attention weights are learned in local windows, and then coarse-grained global self-attention weights are captured by window shift or other operations. In this model, we hypothesize that both fine-grained and coarse-grained self-attentional learning are important. Some recently developed advanced models also support this concept (Hu et al., 2018; Bello et al., 2019; Chen et al., 2019; Srinivas et al., 2021). Experimental results of this paper show that the combination of global and local self-attention can effectively improve performance.

This paper proposed focal self-attention model is shown in Figure 2. The left image shows that feature semantic information is learned by a full self-attention mechanism which will increase the computational complexity by a factor of two. The middle image indicates that global semantic features are captured by a coarse-grained method. The image on the right shows the proposed model combined fine-grained and coarse-grained focal self-attention mechanism. This mechanism divides patch tokens into three levels of granularity. Self-attentional learning operations of different sizes are performed in each window respectively, thus combining local fine-grained and global coarse-grained strategies to capture short-term and remote semantic information more efficiently and effectively.

3. Materials and methods

Focal cross transformer model is a new mechanism to enlarge the receptor field by parallel cross window and improve global dependencies by using local fine-grained and global coarse-grained interactions. The model realizes local and global semantic information interaction by focal self-attention, and uses parallel cross window to enlarge the perceptive field and limit the rapid growth of computational complexity.

3.1. Overall architecture

The overall model utilizes UNet encoder decoder architecture, and the encoder architecture of the Focal cross transformer model is shown in Figure 2. Specifically, the input MRI section of multimodal brain tumor data was formulated by

$$X \in \mathbb{R}^{H \times W \times D \times C} \quad (1)$$

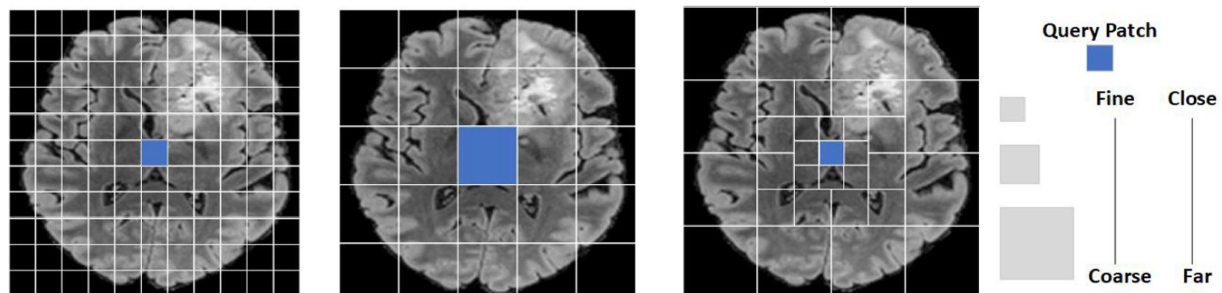


FIGURE 2

A patch token display of a brain tumor input feature map under different granularity levels. The image on the left shows that feature semantic information is learned by a full self-attention mechanism. The intermediate image representation captures the global semantic feature information completely with coarse granularity. The image on the right shows the proposed model combined fine-grained and coarse-grained focal self-attention mechanisms to capture semantic features.

Where the image size is $H \times W \times D$, and the number of input channels of the image is represented by C . Firstly, the image was sliced along the depth direction. For each slice, the input size of the image was formulated by

$$X \in \mathbb{R}^{H \times W \times 4} \quad (2)$$

And then step convolution was used to convert the input image into the patch token of $H/4 \times W/4$. In the encoder path, step convolution was used for down-sampling to acquire the layered architecture. The encoder was divided into four layers; each layer contained N_i focal cross transformers. In the focal cross transformer layer, horizontal and vertical stripes were used for parallel self-attention learning, and fine-grained learning was applied around each token. This paper used coarse-grained strategies at long distances to gain global attention. Next, the feature was transformed by feature mapping; in addition, the image size was halved and the number of channels was doubled by step convolution between layers. Then, we stacked the up-sampling and convolution repeatedly to obtain high-resolution segmentation results.

3.2. Focal cross transformer

Although the original full self-attention mechanism can capture short-term and remote semantic information, its computational complexity is a quadratic form of feature graph size. To alleviate this problem, many researchers tend to use local windows to limit the scope of self-attentional learning, to reduce the computational complexity and memory consumption. Then, the information between local windows is exchanged by shift operation to acquire global information. However, this operation destroys the ability of the original self-attention mechanism to learn both short-term and remote semantic information at the same time. Furthermore, each token can still only obtain semantic information within a limited area; therefore, more blocks need to be stacked to acquire the global receptive field. The focal self-attention based on cross window would enlarge the receptive field and acquire local and global semantic information interactions in a more efficient manner while limiting the rapid growth of computational complexity.

3.3. Focal self-attention

To better realize local and global semantic information interactions, the model used a focal self-attention mechanism that used fine-grained tokens locally and coarse-grained tokens globally, rather than implementing full self-attention mechanism with a fine-grained strategy. Therefore, the global self-attention mechanism can be implemented on the premise of limiting the quadratic increase of computing complexity. Using this system, it was possible to achieve long-term self-attention in less time and with less memory because it only used fine-grained tokens locally and coarse-grained tokens in the long run. However, in practice, we need to query and copy all other tokens for each token, which is still associated with a high computational cost for high-resolution brain tumor images. In view of this, feature mapping was divided into Windows to solve this problem. As shown in Figure 2, the left image represents the use of full self-attention mechanism to learn feature semantic information, which will increase the computational complexity by a factor of two; the middle image represents the use of a coarse-grained strategy to capture global semantic features. However, a large amount of local feature information was lost. The image on the right represents combined fine-grained and coarse-grained focal self-attention mechanism. For the input feature graph By the formula (2), this paper first divided data into a window grid of $S_p \times S_p$, using fine-grained tokens inside the window and coarse-grained tokens outside the window.

To express the proposed method more clearly, this paper defined three terms: feature levels, marked with S_L , which represented the granularity level of extraction for input feature graphs. In Figure 2, this papershow the extraction of three granularity levels. Feature windows size, marked with S_{w_s} , represent the size of the window size in the S_L level and the number of summary tokens, thus providing sub-windows. Feature windows number, marked with S_N , represents the total number of S_w in the S_L tier. By applying these three terms $\{S_L, S_{w_s}, S_N\}$, An module that clearly displays the model, as shown in Figure 2 at the fine-grained level; the three tags are identified $\{3, 11, 11\}$ Figure 3.

3.3.1. Cross window self-attention

As shown in Figure 4, this paper separated the features from fine-grained local self-attention and coarse-grained global self-attention. Taking fine-grained local self-attention as an example, a multi-head self-attention mechanism was used to map the input features to T

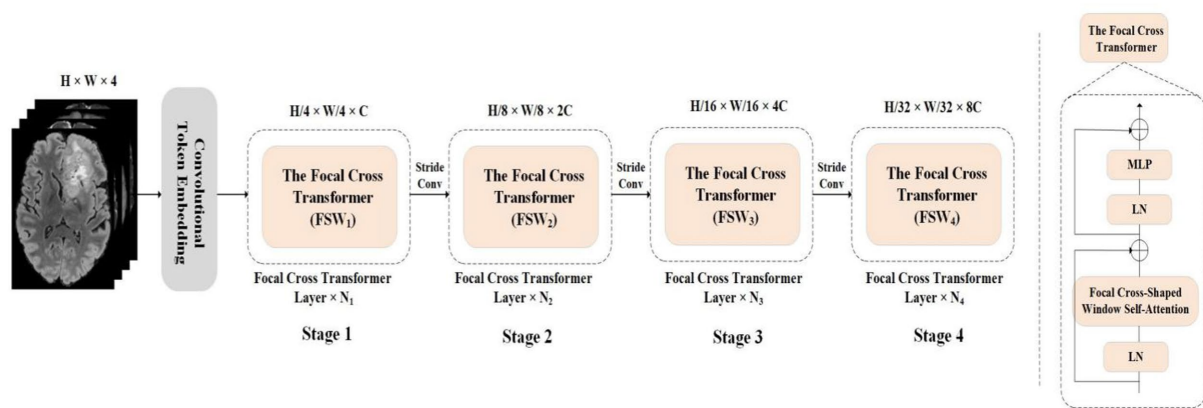


FIGURE 3

The overall architecture of this paper proposed Focal Cross Transformer. The left image represents the encoder path architecture diagram of the overall architecture, and the right image represents the proposed Focal Cross Transformer.

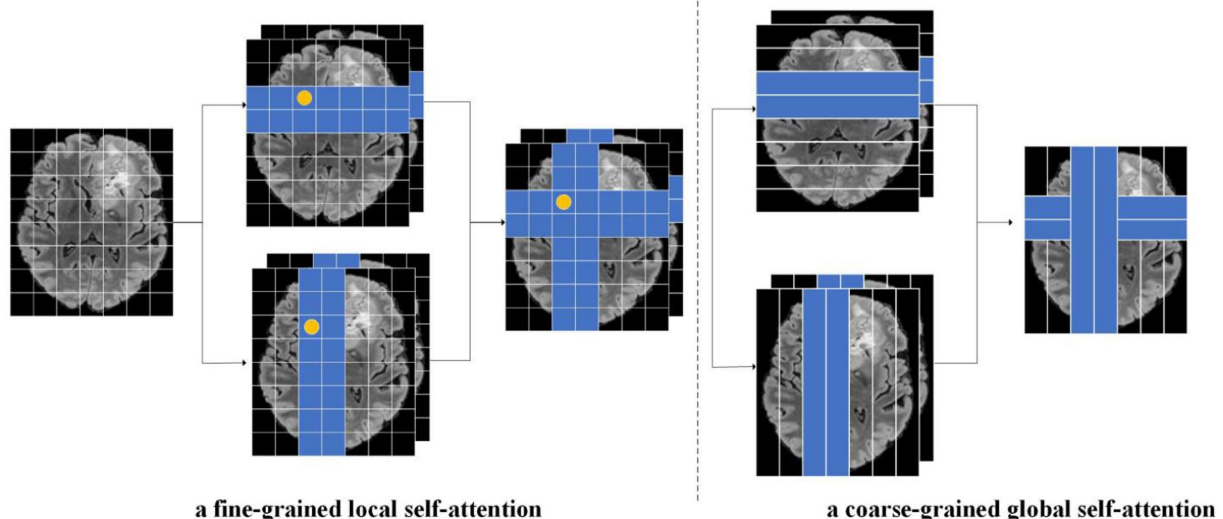


FIGURE 4

As an illustration of this paper focal cross-attention model, the image on the left represents fine-grained local self-attention while the image on the right represents coarse-grained global self-attention. The input feature X was calculated through fine-grained and coarse-grained strategies. In the fine-grained strategy, the input feature was mapped to the T-head and then the head was divided. Next, we calculated horizontal and vertical autogenous attentions in parallel. Finally, the self-attention results for the two parallel groups in the horizontal direction and vertical directions were cascaded. A similar operation was performed for the coarse-grained strategy.

heads; then, each head performed self-attention computations in a horizontal or vertical window (Figure 5).

After mapping the input features to T headers, the headers were segmented to realize parallel computation, where $\{1, 2, \dots, T/2\}$ performs horizontal self-attentional segmentation, $\{T/2 + 1, \dots, T\}$ performs vertical segmentation, and T is usually even. The features were partitioned equally in the horizontal direction and X was partitioned into non-overlapping $[X_1, X_2, \dots, X_M]$ windows of equal width and S_U size. Each window contained $S_U \times W$ tokens. S_U can be used to balance the relationship between self-attention learning and computational complexity, and then fine-grained self-attention weight calculation was carried out for each Token in each $S_U \times W$ size window. Finally, the self-attention results of two parallel groups in horizontal direction and vertical direction were cascaded.

Let us suppose that the dimensions of queries, keys and values of the input feature X projected to the T-th head are all d_i ; then, the formula for calculating self-attention of the T-th head is as follows:

$$X = [X_1, X_2, \dots, X_M] \quad (3)$$

$$Y_t^i = \text{Attention}(X^i W_Q^i, X^i W_K^i, X^i W_V^i) \quad (4)$$

$$\text{Attention}_{H_i}(X) = [Y_t^1, Y_t^2, \dots, Y_t^M] \quad (5)$$

$$X^i \in \mathbb{R}^{(S_U \times W) \times C} \quad (6)$$

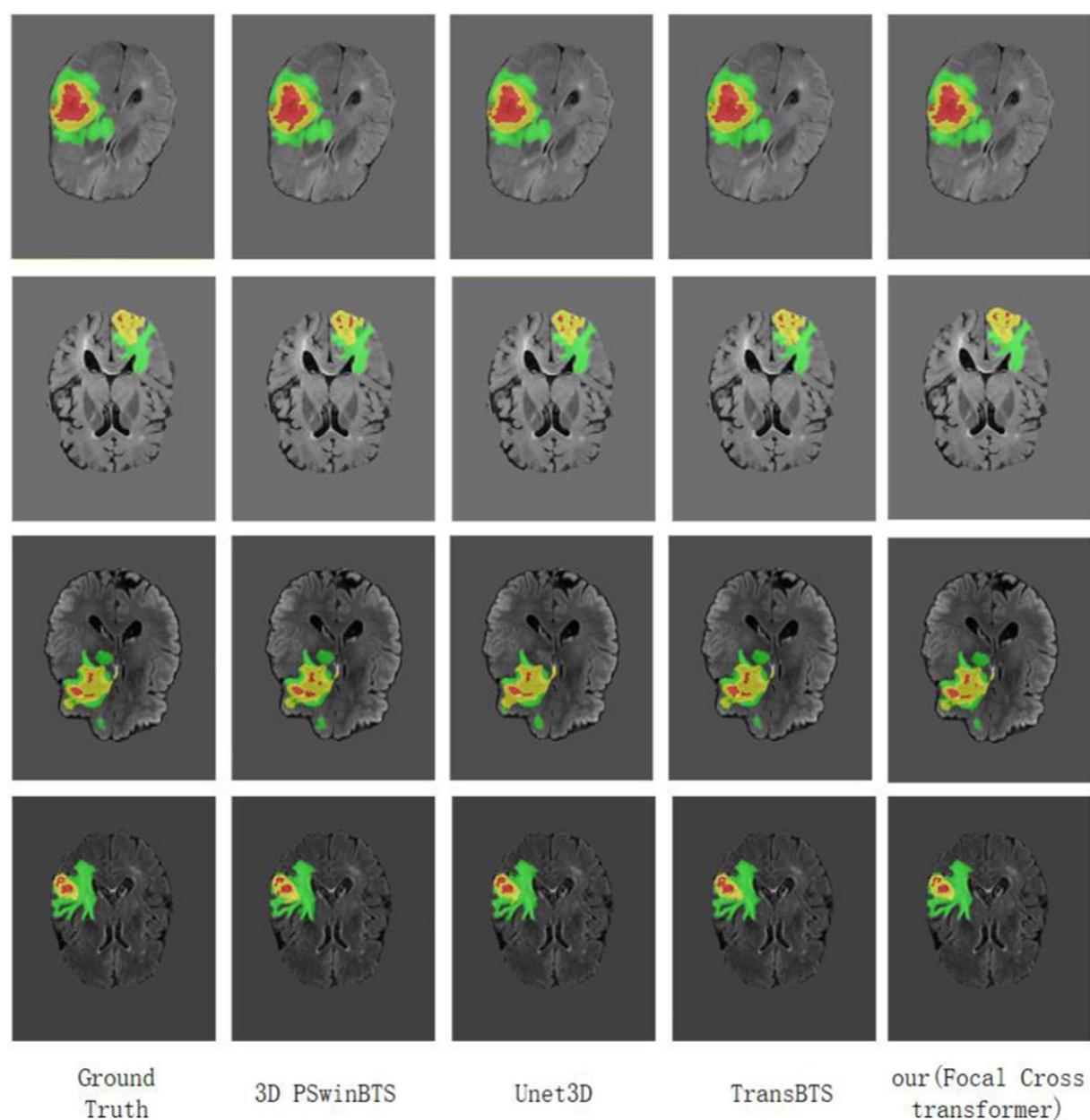


FIGURE 5 Visualization of MRI brain tumor segmentation under different methods. Focal Cross Transformer was compared with the results derived from Unet3D, 3D PSwinBTS, TransBTS, and other models on the BraTS 2021 dataset.

$$M = \frac{H}{S_u}, i = \{1, 2, \dots, M\} \tag{7}$$

$$W_Q^i, W_K^i, W_V^i \in \mathbb{R}^{C \times d}, \tag{8}$$

In these formulae, the corresponding vertical window size is similar. The horizontal and vertical parallel grouping results are then cascaded.

$$\text{Focal Cross - Attention}(X) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_T) \tag{9}$$

3.4. Network encoder

Considering that processing the three-dimensional (3D) Transformer will significantly increase computational complexity and memory consumption, we slice the input feature and slice along the depth direction to obtain a two-dimensional image with input feature

$$X \in \mathbb{R}^{240 \times 240} \tag{10}$$

The overlapped convolutional tokens (kernel = 7, stride 4) were then used to obtain the tokens of

$$\frac{H}{4} \times \frac{W}{4} \left(X \in \mathbb{R}^{60 \times 60} \right) \quad (11)$$

The dimension of each token was C . Then patch token was captured short-term and remote semantic information was acquired through the focal cross transformer layer. In the encoder path, there were four stages, each of which had N_i focal cross transformer layers; this maintained the number of tokens. Each focal cross transformer layer was divided into fine-grained and coarse-grained self-attention mechanisms according to granularity level, thus balancing computational complexity and self-attention learning ability according to granularity. At each level of granularity, the self-attention window range was extended by a parallel Cross window; then, the horizontal and vertical self-attention weights were concatenated. To form a hierarchical structure between the focal cross transformer layers, we used a convolutional layer (kernel=3, stride 2) to reduce the number of tokens and double the channel size. The complete encoder architecture is shown in Figure 3.

3.5. Network decoder

To generate segmentation results in the original slice image, we introduced a CNN decoder for up-sampling and to generate pixel-level segmentation. Slice image features

$$X \in \mathbb{R}^{\frac{H}{32} \times \frac{H}{32} \times 8C} \quad (12)$$

were converted by the feature mapping layer following the encoder layer. Specifically, the sequence data was projected into the standard two-dimensional space through the feature mapping module; then, the image size was expanded and the number of channels was halved by up-sampling through transpose convolution. Then, this paper stacked the upper sampling layer and the convolution layer four times to produce high-resolution segmentation results. Finally, the slices were concatenated to produce segmentation results in the original 3D space.

3.6. Positional encoding

Since the sequence order of the self-attention mechanism remained constant, it can lose important positional information. In an ablation experiment performed previously with Swin Transformer (Liu et al., 2021), it was proven that location information can affect the accuracy of image classification; therefore, researchers tend to use various location coding mechanisms to re-add the lost location information. At present, absolute position coding, relative position coding and conditional position coding are commonly used. The absolute position code uses sinusoidal functions of different frequencies to generate the code, which is then added to the input. Relative position coding considers the distance between markers in the input sequence and can naturally process long sequences of input information during training. Conditional location coding (CPE) relaxes the limitations imposed by explicit location coding on input size, thus allowing Transformer to handle inputs of different sizes.

However, both absolute and relative location coding can add location information to the input token before the Transformer block. This paper concept was derived from the locally enhanced location coding proposed by Dong et al. (2022), in which this model applied location information to the linear projection value and then directly incorporated the location information into each Transformer block.

$$Z_i^t = \sum_{j=1}^n \left(a_{ij}^t + b_{ij}^t \right) v_{ij}^t \quad (13)$$

In Equation (5), Z_i^t represents the T th element of vector Z_i , a_{ij}^t represents the result of calculation at the t th element, the queue, key, and b_{ij}^t represents position coding information. v_{ij}^t represents the value of the self-attention calculation.

4. Experimental results

In this paper, Brats2021 and Brats2019 data sets are used to verify the proposed model. Experimental results and ablation experiments demonstrate that the proposed model extends the receptor field by parallel cross window and improves the global dependence by using local fine-grained and global coarse-grained interactions. It can limit the computational complexity and improve the segmentation accuracy of brain tumors.

4.1. Training data and pre-processing

4.1.1. Training data

The datasets used for model verification in this study were all Brats datasets. This type of dataset is provided by the brain tumor segmentation challenge organized annually by the Medical Image Computing and Computer Assisted Intervention Society (MICCAI). This challenge has been held for 10 consecutive years and exerts significant influence in the field of medical image segmentation. All imaging data sets are manually segmented by 1 to 4 experienced specialists following the same protocol; then, their markings are reviewed by board-certified neuroradiologists. In the present study, the first dataset we used was Brats2021, which included 2,000 patients (8,000 mpMRI scans) including the training set (1,251 patients), the validation set (219 patients), and the test set (530 patients). Each sample consisted of MRI scans from four modes: native T1-weighted (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), T2 Fluid Attenuated Inversion Recovery (T2-flair) volumes, post-contrast T1-weighted (T1GD), T2-weighted (T2), and T2 fluid attenuated inversion recovery (T2-flair) volumes. This paper also included different clinical modalities and a variety of instruments from multiple medical institutions. Each mode had a data size of $240 \times 240 \times 155$ and shared split labels. Each label had four classes {0,1,2,4}: label 0: background; label 1: necrotic tumor core (NCR); label 2: peritumoral edematous/invaded tissue (ED), and label 4: GD-enhancing tumor (ET). The second data set was brats2019, which was not a subset of brats2021; the two datasets were significantly different. The only common data were the images and annotations of BraTS'12-'13; but this did not affect experimental comparisons. The data set included a training set (335 cases) and a validation set (125 cases). The number of samples and modes in each data set were the same.

4.1.2. Pre-processing

All Brats mpMRI scans are available as NIfTI files (.nii.gz). Standardized and enhanced methods were used to process the input data before it was entered into the model for verification. Since the MRI images provided were not standardized, we normalized the gray level of each image and kept the background region as 0. The brats data set has been pretreated with cranial stripping and other procedures. At the same time, four types of data enhancement were implemented in this paper in order to prevent overfitting problems and enhance the Rubon property of the model.

1. Random cropping: considering the large number of black background voxels in the border of the original image, the image was randomly cropped to size $(128 \times 128 \times 128)$ voxels.
2. Random flip: the image is flipped randomly with a probability of 50% along the coronal plane, sagittal plane and axial plane.
3. Intensity normalization: as the data sets are collected from different instruments in different institutions, the image intensity will be different, and it is necessary to carry out intensity normalization. In this paper, Z-Score normalization is used to process data.

$$\bar{X}_j^{(i)} = \frac{X_j^{(i)} - \beta_j}{\alpha_j} \quad (14)$$

Where β is the mean and α is the standard deviation.

4. Gaussian noise: gaussian noise is added to the training process to improve the robustness and generalization ability of the model. Gaussian noise is a noise generated by adding normal distributed random values with a mean of zero and standard deviation to the input data.

4.2. Implementation details and evaluation metrics

4.2.1. Implementation details

This paper trained model with Pytorch, using 8 NVIDIA RTX A5000 (24GB memory) to train 7,050 epochs from scratch using a batch size of 16. For optimization, this paper adopted the Adam optimizer and set its initial learning rate as 0.0003. To achieve more effective convergence, this paper set the decay rate as 0.9 in each iteration. For data set preprocessing, this paper adopted standardization, random flipping, and other strategies to prevent overfitting, but many epochs still needed to be trained. In the training stage, the original training data set was segmented according to a ratio of 8:2 for model training, adjustment, and optimization. According to the inference stage, this paper rescaled the original image and cut the intensity value. Then, this paper uploaded the evaluation model and prediction results to the official website of the host party.

4.2.2. Evaluation metrics

The model used four evaluation metrics for analysis and comparison.

1. The dice similarity coefficient (DSC), which was used to measure the similarity between the brain tumor region

predicted by the proposed Focal Cross transformer and the actual segmentation results provided by Brats; the value range was $[0,1]$ and the greater the value, the higher the accuracy of model prediction. Of these, true positive (TP), the actual brain tumor region, was used to predict the brain tumor region; while true negative (TN) was predicted to be the normal brain tissue region. The false positive (FP) region was actually normal but was predicted to be brain tumor region. The false negative (FN) region was actually negative but was predicted to be normal.

$$\text{Dice} = \frac{2TP}{FP + 2TP + FN} \quad (15)$$

2. Hausdorff_95 (95% HD), the Dice coefficient was sensitive to the region inside the tumor, and the Hausdorff distance was sensitive to the delimited boundary. The Hausdorff_95 represents the last value of the Hausdorff distance multiplied by 95% and was used to eliminate the influence of outlier value small subsets.

$$\text{Hausdorff}_{95} \text{ distance} = P95\{\text{Sup}_{x \in Z} d(x, Y), \text{Sup}_{y \in Y} d(X, y)\} \quad (16)$$

3. Sensitivity, it refers to the proportion of pixels whose true value is tumor that are judged as corresponding tumor or edema.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (17)$$

4. Specificity, it refers to the proportion of pixels that are judged to be normal among the pixels whose true values are normal.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (18)$$

4.3. Main results

4.3.1. Brats 2021

As with previous brain tumor segmentation research, this paper first performed a five-fold cross-validation evaluation on the training set. The average Dice scores of this model for the ET, WT and TC regions were 89.39, 93.58 and 88.65%, respectively. Similarly, at the interface stage, this paper also evaluated the performance of the model by qualitative and quantitative analysis. On the verification set submitted to the official website, we also compared the segmentation results of this model with currently available models; quantitative analysis results are shown in [Table 1](#). The visualized results are shown in [Figure 5](#).

The Dice scores of this model on the BraTS 2021 validation set for ET, TC and WT were 88.28, 86.35 and 93.28% respectively, and the corresponding results of the Hausdorff were 4.58, 5.26 and 3.78, respectively. Compared with a previous classical algorithm ([Table 1](#)), the segmentation accuracy was higher, and the segmentation (in Hausdorff distance) was also significantly improved. Compared with the classical Unet3D model, the Dice coefficient of the model

proposed in this paper for the ET, TC and WT areas, was increased by 9.26, 6.62 and 4.21%, respectively. Since the UNet3D model only used a CNN to learn local feature information, its learning ability for global and long-distance semantic features was insufficient, thus resulting in a big difference between the segmentation accuracy and this model. Compared with the TransBTS model combined with Transformer and UNet, the Dice coefficient of the Focal Cross Transformer method for the ET, TC and WT regions, increased by 1.68, 1.09 and 1.81%, respectively. Compared with the Swin Unter model with layered Swin Transformer, the Dice coefficient of the model proposed in this paper for the ET and WT regions increased by 1.48 and 0.68%, respectively, and decreased by 1.15% in the TC region. In the next experiment, we found that adjusting the width of the stripes in focal cross-attention could further improve the segmentation accuracy of the Focal Cross Transformer model in the TC region, but could lead to a large increase in computational complexity and memory. Therefore, this paper adopted the current configuration on the BraTS 2021 dataset for model validation (Table 2).

4.3.2. Qualitative analysis

This paper visualized the segmentation results of the model on the BraTS 2021 dataset by applying Unet3D, 3D PSwinBTS, TransBTS and other methods. During visual display, we were unable to obtain the ground truth value for the verification set in the BraTS 2021 dataset; thus, this paper performed five-fold cross-validation evaluation of Unet3D, 3D PSwinBTS, TransBTS, and focal cross Transformer model on the training set.

4.3.3. Brats 2019

this paper also evaluated the segmentation results of model on the BraTS 2019 validation set. Because the BraTS 2019 dataset and the BraTS 2021 dataset are different in terms of the number of cases; the sequence type and image size were the same. This paper directly applied hyperparameters on the BraTS 2021 dataset to train model. The average Dice scores of the Focal Cross Transformer model on the BraTS 2019 validation set for ET, WT and TC were 89.68, 93.88 and 89.25%, respectively. The Hausdorff results were 4.32, 4.26 and 3.28, respectively. Compared with the Unet3D, 3D PSwinBTS, and TransBTS models, the Focal Cross Transformer model showed clear improvements in the Dice coefficient and the Hausdorff two evaluation indices (Table 2).

The model presented in this paper achieves excellent performance on BraTS 2019 validation set. This was mainly because the model uses Fine-grained local self-attention and Coarse-grained global self-attention mechanisms to extract the input characteristic information from brain tumors and extract rich local semantic information through fine-grained grained mechanisms. Then, global semantic information was captured with coarse granularity. This strategy effectively improved the pixel level segmentation accuracy.

4.4. Ablation study

To more effectively verify the performance of the model, this paper performed extensive ablation experiments to prove the rationality and feasibility of the model's design principle. This paper investigated the model's capabilities in several different ways. Unet3D,

TABLE 1 Comparison and analysis of the BraTS 2021 validation set.

Method	Enhancing tumor (ET)				Tumor core (TC)				Whole tumor (WT)			
	Dice (%)	HD95 (mm)	Sensitivity (%)	Specificity (%)	Dice (%)	HD95 (mm)	Sensitivity (%)	Specificity (%)	Dice (%)	HD95 (mm)	Sensitivity (%)	Specificity (%)
Unet3D (Akbar et al., 2022)	78.02	25.82	80.51	99.97	80.73	21.17	80.55	99.97	89.07	11.78	92.34	99.88
Multi-scale features (Li et al., 2021)	76.89	30.21	78.69	99.97	81.61	16.65	80.57	99.96	90.18	6.16	88.33	99.91
Swin unter (Hatamizadeh et al., 2022)	85.8	6.02	83.68	99.96	88.5	3.77	86.74	99.98	92.6	5.83	93.65	99.95
Evaluating scale attention (Yuan, 2021)	84.79	12.75	-	-	86.55	11.19	-	-	92.65	3.67	-	-
TransBTS	85.16	19.26	83.14	99.97	86.26	12.38	85.74	99.95	91.47	10.62	93.61	99.90
3D PSwinBTS (Liang et al., 2022)	79.48	19.44	81.31	99.95	84.20	7.25	85.11	99.97	90.76	5.57	92.59	99.94
Our (Focal cross transformer)	87.28	4.58	85.62	99.98	87.35	5.26	86.89	99.99	93.28	3.78	94.22	99.97

TABLE 2 Comparison and analysis of the BraTS 2019 validation set.

Method	Enhancing tumor (ET)				Tumor core (TC)				Whole tumor (WT)			
	Dice (%)	HD95 (mm)	Sensitivity (%)	Specificity (%)	Dice (%)	HD95 (mm)	Sensitivity (%)	Specificity (%)	Dice (%)	HD95 (mm)	Sensitivity (%)	Specificity (%)
Unet3D	83.26	23.18	81.32	99.95	82.32	22.28	80.48	99.96	89.58	14.24	91.68	99.88
Swin uniter	86.27	7.98	83.39	99.97	89.23	4.68	85.07	99.96	91.23	8.36	92.09	99.91
TransBTS	84.52	17.23	78.69	99.95	85.18	11.27	85.66	99.94	90.21	11.42	92.78	99.89
3D PSwinBTS	81.71	15.63	82.68	99.96	82.46	9.63	84.68	99.97	87.62	7.92	91.85	99.93
Our (Focal cross transformer)	89.68	4.32	85.38	99.97	89.25	3.28	85.96	99.98	93.88	4.26	93.85	99.95

TABLE 3 Ablation study on coarse-grained global and fine-grained local mechanism.

Method	Dice (%)		
	ET	TC	WT
Coarse-grained	85.26	84.32	89.59
Coarse-grained global and fine-grained local	87.28	87.35	93.28

3D PSwinBTS and TransBTS proved that the combination of CNN and Transformer effectively improved the performance of the model. Therefore, this paper no longer independently verified the influence of CNN and Transformer on the performance for brain tumor segmentation.

4.4.1. Coarse-grained global and fine-grained local

This paper used fine-grained tokens locally and coarse-grained tokens globally, rather than implementing a full self-attention fine-grained mechanism. The combination of coarse-grained global self-attention and fine-grained local attention mechanism is an important aspect of the model proposed in this paper. However, full self-attention adopted by vision Transformer cannot be applied to brain tumor segmentation due to high levels of computational complexity. Therefore, it is not possible to verify cases that only use fine-grained full self-attention mechanisms. This paper only verified the comparative performance between a model that adopted the combination of global coarse-grained and local fine-grained mechanisms and a model with the same granularity. This paper use the combined CNN and cross Transformer model in the encoder to perform a comparison experiment between the segmentation of brain tumors with the same particle size and the current model combined with coarse-grained global and fine-grained local mechanisms. The input features size is shown in Formula (1); then, slices were generated along the depth direction. For each slice and the input size of the image is shown in Formula (2), step convolution was used to convert the input image into a patch token of $H/4 \times W/4$. In the encoder path, step convolution was used for down-sampling to achieve the layered architecture. Table 3 shows the results of comparative experiments. For ET, TC and WT, Dice coefficients of the coarse-grained global and fine-grained local models increased by 2.02, 3.03 and 3.69%, respectively.

4.4.2. Cross window

In the model, this paper extended the scope of the self-attention window by applying a parallel cross window and then concatenated the horizontal and vertical self-attention weights. This paper created $sw = 1$ and $sw = 2$ Windows separately in the horizontal direction to learn self-attention, and the same configuration was also adopted in the vertical direction; ‘sw’ indicates the size of the sharded self-attention window width. Table 4 shows the Dice coefficients of self-attentional learning and cross window model for ET, TC and WT in the horizontal and vertical directions, respectively. By performing comparative experiments, this paper proved that by combining horizontal and vertical self-attention weights, this model effectively increased the receptive field of the self-attention window and improved the segmentation performance of the model.

TABLE 4 Ablation study on cross window.

Method	Dice (%)		
	ET	TC	WT
Horizontal (sw = 1)	79.38	81.24	83.62
Horizontal (sw = 2)	84.62	85.74	88.49
Vertical (sw = 1)	80.02	82.39	86.27
Vertical (sw = 2)	84.76	86.95	87.83
Cross window	87.28	87.35	93.28

5. Conclusion

This paper developed a novel segmentation model for brain tumors. Fine-grained local self-attention and coarse-grained global self-attention mechanisms were combined to extract characteristic input information from brain tumors and extract rich local semantic information through fine-grained mechanisms. Then, global semantic information was captured with coarse granularity. The cross window concurrent multi-head and self-attention mechanism was used to learn the self-attention weight in the window from both horizontal and vertical directions, thus expanding the receptive field of self-attention learning. This also balanced the relationship between computational complexity and self-attention learning ability in Transformer. Experimental results on the Brats2021 and Brats2019 datasets validated proposed model. In future research, we will continue to explore ways to improve Transformer’s global self-attention learning ability and reduce computational complexity so that we can build an efficacious segmentation model for brain tumors.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://www.med.upenn.edu/cbica/brats2021/>.

References

Akbar, A. S., Fatichah, C., and Suciati, N. (2022). “Unet3D with Multiple Atrous Convolutions Attention Block for Brain Tumor Segmentation”, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, BrainLes 2021. Lecture Notes in Computer Science*, vol 12962. eds. Crimi, A., Bakas, S. Springer, Cham. doi: 10.1007/978-3-031-08999-2_14

Bello, I., Zoph, B., Vaswani, A., Shlens, Jonathon, and Le, Quoc V. (2019). “Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation”, in *Computer Vision – ECCV 2022 Workshops. ECCV 2022. Lecture Notes in Computer Science*, eds. Karlinsky, L., Michaeli, T., Nishino, K. vol 13803. Springer, Cham.

Cao, H., Wang, Y., Chen, J., Jiang, Dongsheng, Zhang, Xiaopeng, Tian, Qi, et al. (2023). “Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation”, in *Computer Vision – ECCV 2022 Workshops. ECCV 2022. Lecture Notes in Computer Science*, vol 13803 eds. Karlinsky, L., Michaeli, T., Nishino, K. (eds). Springer, Cham.

Chen, Jieneng, Lu, Yongyi, Yu, Qihang, Luo, Xiangde, Adeli, Ehsan, Wang, Yan, et al. (2019). GCNet: non-local networks meet squeeze-excitation networks and beyond.2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 1971–1980.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., et al. (2021). Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*. doi: 10.48550/arXiv.2102.04306

Chu, X., Tian, Z., Wang, Y., Zhang, B., Ren, H., Wei, X., et al. (2021a). Twins: revisiting the design of spatial attention in vision transformers. *Adv. Neural Inf. Proces. Syst.* 34, 9355–9366.

Author contributions

LZ wrote the main content of the manuscript and carried out experimental research. WS edited and supervised main content of the manuscript. FS and YG put forward suggestions on the structure and experimental part of the paper, and verified by experiments. All the authors reviewed the manuscript and agreed to publish it.

Funding

This study was supported by Analysis, Prediction and Intervention of Complex Network Behavior in Multilingual Big Data environment, 61433012, National Natural Science Foundation of China, National Key Research and Development Program of Internet Chinese Information Processing and Verification System for Public Security and Social Management, 2014CB340506, automatic segmentation system of brain tumor based on Information Security Technology, 22JR11RA004, Gansu Youth Science and Technology Foundation Program.

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.

Chu, X., Tian, Z., Zhang, B., Wang, X., Wei, X., Xia, H., et al. (2021b). Conditional positional encodings for vision transformers. *arXiv preprint arXiv:2102.10882*. doi: 10.48550/arXiv.2102.10882

Dong, X., Bao, J., Chen, D., Zhang, Weiming, Yu, Nenghai, Yuan, Lu, et al. CSWin Transformer: A General Vision Transformer Backbone with Cross-Shaped Windows. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (2022) 12114–12124. doi: 10.48550/arXiv.2107.00652

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. doi: 10.48550/arXiv.2010.11929

Gao, L., Zhang, J., Yang, C., and Zhou, Y. (2022). Cas-VSwin transformer: a variant swin transformer for surface-defect detection. *Comput. Ind.* 140:103689. doi: 10.1016/j.compind.2022.103689

Hatamizadeh, A., Nath, V., Tang, Y., Yang, Dong, Roth, Holger, and Xu, Daguang (2022). Swin UNETR: swin transformers for semantic segmentation of brain tumors in mri images. *Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries: 7th international workshop, BrainLes 2021, held in conjunction with MICCAI 2021, virtual event, September 27, 2021, revised selected papers, part I*, Cham: Springer International Publishing 272–284.

He, S., Luo, H., Wang, P., Wang, F., Li, H., and Jiang, W. (2021). TransReID: transformer-based object re-identification. *TransReID: Transformer-based Object Re-Identification. 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 14993–15022. doi: 10.48550/arXiv.2102.04378

- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). MobileNets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. doi: 10.48550/arXiv.1704.04861
- Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 7132–7141. doi: 10.48550/arXiv.1709.01507
- Huang, J., Fang, Y., Wu, Y., Wu, H., Gao, Z., Li, Y., et al. (2022). Swin transformer for fast MRI. *Neurocomputing* 493, 281–304. doi: 10.1016/j.neucom.2022.04.051
- Li, Z., Shen, Z., Wen, J., He, Tian, and Pan, Lin (2022). Automatic brain tumor segmentation using multi-scale features and attention mechanism. Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 7th international workshop, BrainLes 2021, held in conjunction with MICCAI 2021, virtual event, September 27, 2021, revised selected papers, part I. Cham: Springer International Publishing, 216–226.
- Liang, J., Cao, J., Sun, G., Zhang, Kai, Van Gool, Luc, and Timofte, Radu (2021). SwinIR: image restoration using swin transformer. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 1833–1844. doi: 10.48550/arXiv.2108.10257
- Liang, J., Yang, C., and Zeng, L. (2022). 3D PSwinBTS: an efficient transformer-based Unet using 3D parallel shifted windows for brain tumor segmentation. *Digit. Signal Process.* 131:103784. doi: 10.1016/j.dsp.2022.103784
- Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G., and Zhang, D. (2022). Ds-TransUNet: dual swin transformer U-net for medical image segmentation. *IEEE Trans. Instrum. Meas.* 71, 1–15. doi: 10.1109/TIM.2022.3178991
- Liu, Z., Hu, H., Lin, Y., Yao, Zhuliang, Xie, Zhenda, and Wei, Yixuan (2022). Swin transformer v2: scaling up capacity and resolution. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 12009–12019. doi: 10.48550/arXiv.2111.09883
- Liu, Z., Lin, Y., Cao, Y., Hu, Han, Wei, Yixuan, Zhang, Zheng, et al. (2021). Swin transformer: hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF international conference on computer vision, 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 9992–10002. doi: 10.48550/arXiv.2103.14030
- Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3431–3440. doi: 10.48550/arXiv.1411.4038
- Srinivas, A., Lin, T. Y., Parmar, N., Shlens, Jonathon, Abbeel, Pieter, and Vaswani, Ashish (2021). Bottleneck transformers for visual recognition. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 16519–16529.
- Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., and Jegou, H. (2020). Training data-efficient image transformers & distillation through attention. *ArXiv*, abs/2012.12877. doi: 10.48550/arXiv.2012.12877
- 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, 600–6010. doi: 10.48550/arXiv.1706.03762
- Wang, W., Chen, C., Ding, M., Li, Jiangyun, Yu, Hong, and Zha, Sen (2021a). TransBTS: Multimodal Brain Tumor Segmentation Using Transformer. In: et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021 MICCAI 2021. Lecture Notes in Computer Science, vol 12901. Springer, Cham. 109–119.
- Wang, W., Xie, E., Li, X., Fan, Deng-Ping, Song, Kaitao, Liang, Ding, et al. (2021b). Pyramid vision transformer: a versatile backbone for dense prediction without convolutions. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 548–558. doi: 10.48550/arXiv.2102.12122
- Yuan, Y. (2021). Evaluating scale attention network for automatic brain tumor segmentation with large multi-parametric MRI database[C]. Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries: 7th international workshop, BrainLes 2021, held in conjunction with MICCAI 2021, virtual event, September 27, 2021, revised selected papers, part II Cham: Springer International Publishing, 2022: 42–53.
- Yuan, L., Chen, Y., Wang, T., Yu, Weihao, Shi, Yujun, Jiang, Zihang, et al. (2021). Tokens-to-token vit: training vision transformers from scratch on imagenet. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 538–547. doi: 10.48550/arXiv.2101.11986
- Zhang, P., Dai, X., Yang, J., Xiao, Bin, Yuan, Lu, Zhang, Lei, et al. (2021). Multi-scale vision longformer: a new vision transformer for high-resolution image encoding. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2978–2988. doi: 10.48550/arXiv.2103.15358
- Zheng, S., Lu, J., Zhao, H., Zhu, Xiatian, Luo, Zekun, Wang, Yabiao, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2021) 6877–6886. doi: 10.48550/arXiv.2012.15840
- Zhu, X., Su, W., Lu, L., Li, Bin, Wang, Xiaogang, and Dai, Jifeng (2021). DD deformable transformers for end-to-end object detection. Proceedings of the 9th International conference on learning representations, virtual event, Austria. 3–7.



OPEN ACCESS

EDITED BY

Shuqiang Wang,
Chinese Academy of Sciences (CAS), China

REVIEWED BY

Xuhang Chen,
University of Macau, China
Ximei Zhu,
Peking University, China

*CORRESPONDENCE

Jianfeng He
✉ jfenghe@foxmail.com

RECEIVED 26 February 2023

ACCEPTED 08 May 2023

PUBLISHED 24 May 2023

CITATION

Xiong X, Feng J, Zhang Y, Wu D, Yi S, Wang C,
Liu R and He J (2023) Improved
HHT-microstate analysis of EEG in nicotine
addicts.
Front. Neurosci. 17:1174399.
doi: 10.3389/fnins.2023.1174399

COPYRIGHT

© 2023 Xiong, Feng, Zhang, Wu, Yi, Wang, Liu
and He. 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.

Improved HHT-microstate analysis of EEG in nicotine addicts

Xin Xiong¹, Jiannan Feng¹, Yaru Zhang¹, Di Wu¹, Sanli Yi¹,
Chunwu Wang², Ruixiang Liu³ and Jianfeng He^{1*}

¹Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China, ²College of Physics and Electronic Engineering, Hanshan Normal University, Chaozhou, China, ³Department of Clinical Psychology, Second People's Hospital of Yunnan, Kunming, China

Background: Substance addiction is a chronic disease which causes great harm to modern society and individuals. At present, many studies have applied EEG analysis methods to the substance addiction detection and treatment. As a tool to describe the spatio-temporal dynamic characteristics of large-scale electrophysiological data, EEG microstate analysis has been widely used, which is an effective method to study the relationship between EEG electrodynamics and cognition or disease.

Methods: To study the difference of EEG microstate parameters of nicotine addicts at each frequency band, we combine an improved Hilbert Huang Transformation (HHT) decomposition with microstate analysis, which is applied to the EEG of nicotine addicts.

Results: After using improved HHT-Microstate method, we notice that there is significant difference in EEG microstates of nicotine addicts between viewing smoke pictures group (smoke) and viewing neutral pictures group (neutral). Firstly, there is a significant difference in EEG microstates at full-frequency band between smoke and neutral group. Compared with the FIR-Microstate method, the similarity index of microstate topographic maps at alpha and beta bands had significant differences between smoke and neutral group. Secondly, we find significant class × group interactions for microstate parameters at delta, alpha and beta bands. Finally, the microstate parameters at delta, alpha and beta bands obtained by the improved HHT-microstate analysis method are selected as features for classification and detection under the Gaussian kernel support vector machine. The highest accuracy is 92% sensitivity is 94% and specificity is 91%, which can more effectively detect and identify addiction diseases than FIR-Microstate and FIR-Riemann methods.

Conclusion: Thus, the improved HHT-Microstate analysis method can effectively identify substance addiction diseases and provide new ideas and insights for the brain research of nicotine addiction.

KEYWORDS

addictions, improved HHT-microstate, EEG, frequency band, detect

1. Introduction

Substance addiction is a chronic relapsing disease, which refers to the adaptation and dependence of individuals after long-term abuse of harmful substances ([World Health Organization, 2009](#)). It includes drug addiction and other mental addictions such as alcohol, nicotine, and caffeine. The root cause is the long-term adaptation of the brain to addictive substances, which makes it difficult for individuals to give up due to the escalating behavior of substance intake, even though they have been aware of the negative effects ([Pengfei et al., 2019](#)). Substance addiction can lead to a range of diseases, such as lung cancer, iron-deficiency heart

disease and esophageal cancer, and cause an enormous emotional, financial and medical burdens on individuals and society. Previous studies on the cognitive function of addicts with different substances have shown that addiction have an impact on cognition. For example, addicts have impaired executive control function, increased impulsivity (Fulton and Charlotte, 2009), decreased decision-making ability (Zernig et al., 2007), and strong memories related to addiction cues (Robbins et al., 2008; Yan Xue et al., 2017); Meanwhile, drug addicts have abnormal sleep structure (Conroy and Arnedt, 2014), whose sleep stages affected by various addictive substances (Colrain et al., 2014). In addition, addiction theory has also shown that a common character of substance addiction is drug cue response, which means that compared with non-addicts, addicts show significantly different physiological and psychological responses when they face with smoke cues or neutral cues (Linyuan and Xiaoyi, 2005). Moreover, the responsiveness of addicts to cigarette-related cues is also the main factor of relapse (Wei et al., 2017), which means that the brain response of addicts to cigarette cues may predict their ability to give up smoking continually. At present, the main treatment methods of substance addiction are drug therapy, psychotherapy, physical therapy, and neutral feedback therapy. In recent years, the technique of Brain Computer Interface (BCI) from the perspective of electrophysiology has been proposed and applied to the research of addiction. BCI is a communication system built between the brain and other external devices, rather than relying on the brain transmission pathway composed of peripheral nerves and muscles, which is a new way of human-computer interaction. By using a non-invasive, cheap and powerful tool, Electroencephalogram (EEG), it can record the configuration of brain electric fields produced by the coordination of different nerve combinations and have a high temporal resolution (Arshad et al., 2022; Totev et al., 2023). Besides, with the development of the computer technology, some relevant and novel algorithms including Generative Adversarial Network (GAN), Convolutional Neural Networks (CNN) were employed for features extraction at EEG bands or to explore the potential neural mechanism of the brain (Hu et al., 2020; Prasanth et al., 2020; Yan et al., 2020), and also made great contributions in some specific states or diseases such as Parkinson, depression and epilepsy (Uyulan et al., 2020; Chu et al., 2021; Gabeff et al., 2021). On this basis, studies showed that the brain of substance addict has abnormal functions and structural changes (Bjork and Gilman, 2014). Therefore, many researchers have processed the cognitive function of the brain in different states and collected signals from the cerebral cortex to analyze the mechanism of addiction.

Previous investigations and studies on EEG signals of addicts have shown that there are some qualitative and quantitative changes in EEG signals of these addicts, including EEG coherence, frequency domain features and nonlinear features, and EEG source localization. As a measurement of brain network, coherence reflects functional connectivity and activity synchronization among brain regions (Franken et al., 2004) and has advantages in terms of high temporal resolution and measurement of brain networks among neuron populations. Compared with non-addicted people, EEG coherence in addicted people is significantly enhanced (Yan Xue et al., 2017) and significantly correlate with the changes in smoking cravings (Littel et al., 2009), which is advisable to explore the changes in brain activity related to addiction. In the frequency domain, there are less alpha EEG rhythm and more beta EEG rhythm in the addicted people, and

many delta-theta rhythms with low amplitude in the central brain region (Benos and Kapinas, 1980; Olivennes et al., 1983; Gekht et al., 2002); addicts have higher EEG correlation dimension than non-addicts, which can reflect the attention deficit of addicted people (Sv  tl  k et al., 2010). Furthermore, the analysis method of EEG source localization has also been applied to research the mechanism of addiction and solved the problem of observing the difference of brain activity and locating deep source error in high temporal and spatial resolution (Pascualmarqui et al., 1994). In addition, the rank-based feature selection method was used to assign weight values to EEG features such as the interhemispheric coherence and spectral power at EEG bands of patients with alcohol disorders, which obtain the better accuracy with the classification of the most discriminative features (Mumtaz et al., 2017). Meanwhile, study also found that the theta band (4–8 Hz) between the frontal and posterior cortical regions had a high level of synchronization in the brain of drug addicts according to the connectivity of subband cortical network which was calculated by synchronization likelihood algorithm (Coullaut-Valera et al., 2014).

EEG microstate is one of the methods to determine and quantify the oscillatory activity and dynamic characteristics of the cerebral cortex. It was first proposed by Lehmann, who regarded the multi-channel spontaneous EEG signals as a series of EEG topographic maps changing over time. Each EEG topographic map is the superposition of the effects of all the sources that are instantly active at present and is a global measure of instantaneous EEG activity (Lehmann et al., 1987; Lehmann, 1994). It reveals that the distribution of brain electrical activity does not change continuously but discretely over time. The topological structure of one EEG topographic map does not smoothly change to another structure, but stays in a quasi-stable state for about 80–120 ms, and then suddenly changes to another structure. Several EEG topographic maps with the same topological structure are classified as a class of microstate (Arjun et al., 2014). In the literature of microstate analysis, generally four different microstate, typically labeled from A to D, respectively correspond to the activities of auditory network, visual network, prominence network and attention network (Britz et al., 2010) and can usually explain more than 80% of the variance present in the EEG data (Lehmann et al., 2005; Seitzman et al., 2017). The temporal parameters of microstate include global interpretation variance, mean duration, occurrence and coverage of microstates, which provide new avenues for quantifying cortical oscillatory activity with functional relevance. The changes of these parameters can reflect the impact of diseases on the brain, such as Parkinson's disease (John et al., 2023), dementia (Grieder et al., 2016), schizophrenia (Andreou et al., 2014) and Alzheimer's disease (Strik et al., 1997). Besides, there are also studies on the identification of epilepsy (Kiran et al., 2018) and motor imagination (Weifeng et al., 2017) by using microstate parameters, and the accuracy is relatively high. However, most previous studies on EEG of addicted subjects were based on the analysis of brain network and EEG characteristics. Therefore, using microstate analysis to compare different EEG of addicts is a valuable method to analysis and detect substance addiction.

Previous studies have shown that the dynamics of microstates in the time domain are correlated with those in the spectrum domain. Milz et al. (2017) found that there was a consistent relationship between intra-band microstates and power, which meant that the intra-cortical intensity and spatial distribution of alpha frequency band were determined. Traditional microstate analysis method lacks frequency domain information (Koenig et al., 2018), which affects the conclusion of

correlation between microstate and spectrum domain. In order to solve this problem, Ehtasham et al. (2019) used the empirical mode decomposition (EMD) and instantaneous frequency model in the Hilbert Huang Transform (HHT) method to extract the spectral features of microstates in time series. This method can preserve the local spectral properties of the original data in time domain, and does not require prior characteristic information of data, or as in the case of other decomposition methods such as Fourier, or wavelet analysis, it does not assume a pre-determined set of basis functions (Daubechies et al., 2011; Thakur et al., 2013). Therefore, in this paper, an improved HHT-Microstate method was used to research the EEG of nicotine addicts. By preserving the instantaneous properties of the data in the spectrum domain, the microstate time series was analyzed to evaluate the instantaneous changes of the spectral features of the EEG data.

In this study, we selected two different types of nicotine addiction EEG data and used the improved HHT method to divide the EEG data into five frequency bands, including delta band (0.5–4 Hz), theta band (4–8 Hz), alpha band (8–12 Hz), beta band (12–30 Hz), and gamma band (30–40 Hz). By comparing the differences of the microstate parameters between the two types of tasks at each frequency band, we selected the microstate parameters with significant differences as features to detect different types of nicotine addiction. At the same time, in order to highlight the superiority of the improved HHT-Microstate method, the frequency band division method and EEG feature analysis method are compared with Finite Impulse Response (FIR) method and EEG Riemann distance method, which includes FIR-Microstate method, HHT-Riemann method and FIR-Riemann method. Finally, we proved that the improved HHT-Microstate method is superior to other methods and can detect and identify different addiction states more effectively.

2. Materials and methods

2.1. Method

Our experiments in this paper are carried out on the platform of Matlab_2019 and the corresponding toolbox of EEGLAB_2019. The main method has two parts. Firstly, the data is time-frequency decomposition by using the Empirical Mode Decomposition (EMD) and instantaneous frequency model. Secondly, microstate analysis method is applied to each frequency band for the extraction of microstate topographic maps and microstate parameters at each frequency band.

2.1.1. Time-frequency analysis—improved Hilbert Huang transform

The obstacle to finding a correlation between microstates and spectrum is to correlate microstates in different temporal resolutions with spectral analysis. On the one hand, the EEG microstate analysis is carried out in the time domain and determine the EEG data and similarity index of the given segmentation for each instance; On the other hand, traditional spectrum analysis methods require at least a period to calculate the spectral power of any given frequency band. Therefore, in order to solve this obstacle, the EMD and instantaneous frequency models in HHT (Huang, 1998) were used for time-frequency analysis (Huang et al., 2009; Liu et al., 2022). However, in the process of decomposition, the traditional EMD will cause

problems of mode aliasing, which makes the component lose the single feature scale feature. Ensemble Empirical Mode Decomposition (EEMD) model has been optimized for EMD, but the added Gaussian white noise will remain and affect the result. To solve the problem of mode aliasing in EMD and reconstruction error or low computational efficiency in EEMD, a CEEMDAN method is used in this paper. By adding adaptive white noise to the EEMD, errors in signal reconstruction can be reduced and high computational efficiency can be ensured. Besides, it could maintain the original temporal resolution while transforming time-domain data into time-frequency domain data. The main steps of CEEMDAN are as follows:

- (1) Signal can obtain several Intrinsic Mode Functions (IMFs) after EMD, and each IMF must satisfy two restrictions: ① The difference between the number of extreme points and zero points is not more than 1; ② The mean value between the local maximum envelope and the local minimum envelope at any time is 0.

Add white noise $\beta A_j(t)$ to the original signal $x(t)$, where β is the standard noise and j is the number of noises added. The newly constructed signal is $Z(t) = x(t) + \beta A_j(t)$. The first order component of CEEMDAN is:

$$IMF_1 = \frac{1}{n} \sum_{j=1}^n IMF_1^j \quad (1)$$

The remainder is $r_1 = Z(t) - IMF_1$.

- (2) The original signal of the second IMF component is $Z(t) = r_1 + \beta A_j(t)$, after decomposition, it can obtain:

$$IMF_2 = \frac{1}{2n} \sum_{j=1}^{2n} IMF_2^j \quad (2)$$

The remainder is $r_2 = Z(t) - IMF_2$.

- (3) Next, repeat step (1) and (2) until the decomposition is complete. m IMF components are obtained, and the residual is:

$$r = x(t) - \sum_{i=1}^m IMF_i \quad (3)$$

- (4) The reconstruction formula is:

$$Z(t) = r + \sum_{i=1}^m IMF_i \quad (4)$$

Due to the existence of false IMF components in the process of EMD decomposition, these false IMF components should be eliminated in practical application. The correlation between the real IMF component and the original signal is greater than those in false component, and the proportion of the real IMF component is larger than those in false IMF component. At present, the commonly used methods to eliminate false IMF components include correlation coefficient method, gray correlation degree method, mutual Information method, energy ratio method and K-S test method. Gray correlation degree method and K-S test method can better distinguish the false IMF component for single component signals, but it is difficult to distinguish the complex signals with multiple components (Yang et al., 2013). Correlation coefficient method and energy ratio method

have great amplitude dependence on signals, which is not conducive to the differentiation of false IMF components (Bao et al., 2009). Mutual Information (represented by the symbol IMI) can accurately calculate the correlation between the IMF component and the original signal, and has certain advantages in distinguishing the false IMF component.

Therefore, our paper uses IMI to select IMF components. IMI describes the degree of correlation between two random variables, and the amount of common information between two variables can be measured by IMI. The larger the IMI, the more common information between variables, and the stronger the correlation. For the i th IMF component $c_i(t)$ of the signal and the original signal $x(t)$, the IMI between them is defined as:

$$I_{MI}(c_i; x) = -\sum p(c_i; x) \log \frac{p(c_i; x)}{p(c_i)p(x)} \quad (5)$$

where, $p(c_i)$ and $p(x)$ are the marginal probability distributions of the i th IMF component $c_i(t)$ and the original signal $x(t)$ respectively; $p(c_i, x)$ is the joint probability distribution of the i th IMF component $c_i(t)$ and the original signal $x(t)$.

After the decomposition of CEEMDAN and the selection of IMI, HT is used to calculate the instantaneous frequency and amplitude for IMF. The impulse response of HT is:

$$h(t) = \frac{1}{\pi t} \quad (6)$$

The HT expression of IMF is:

$$H(IMF(t)) = h(t) * IMF(t) \quad (7)$$

where, $H(\bullet)$ is the function of HT, $*$ is the convolution. Then:

$$Z(t) = IMF(t) + jH(IMF(t)) = \alpha(t)e^{j\theta(t)} \quad (8)$$

where

$$\alpha(t) = \sqrt{IMF^2(t) + H^2(IMF(t))} \quad (9)$$

$$\theta(t) = \arctan \frac{H(IMF(t))}{IMF(t)} \quad (10)$$

Therefore, the instantaneous frequency can be expressed as:

$$F(t) = \frac{d\theta(t)}{dt} \quad (11)$$

$F(t)$ and $\alpha(t)$ are the instantaneous frequency and amplitude of IMF, respectively. Based on the instantaneous frequency value of IMF, IMFs with different instantaneous frequency values can be obtained by selecting different sampling frequency for each electrode data and decomposing them. The instantaneous frequency value is adjusted to the frequency range of the above 5 frequency bands, which can obtain the corresponding 5 frequency bands. In addition, the time information can be saved after microstate extraction.

2.1.2. Microstate analysis

The core of microstate analysis is to segment EEG data into microstates by using clustering algorithms. The well-established standardized procedures in EEGLAB (Poulsen et al., 2020) are used for the microstate analysis. The specific processes are as follows:

- (1) The quantized scalar values for electric potentials across EEG electrodes also known as Global Field Potential (GFP) are computed for EEG: the standard deviation of voltage values at all electrodes of a topographic map at a time, which is used to describe the strength of the electric field of a topographic map. The formula is as follows:

$$GFP(t) = \sqrt{\frac{1}{K} \left(\sum_{i=1}^K [V_i(t) - V_{mean}(t)]^2 \right)} \quad (12)$$

where, K is the number of channels, $V_i(t)$ is the voltage and potential at the i th electrode, and $V_{mean}(t)$ is the instantaneous average potential between electrodes.

GFP represents the intensity of the electric field on the brain at every moment. It is usually used to measure the total response of the brain to the event or to represent the rapid changes in brain activity. The local maximum of its curve represents the moment of the strongest field intensity and the highest signal-to-noise ratio. Therefore, using the topographic map at the peak of GFP to represent other surrounding topographic maps for analysis is an effective method to improve the microstate signal-to-noise ratio and reduce the amount of computation (Murray et al., 2008). At the same time, it is also found that the topographic map at the peak of GFP is similar to the surrounding one, while the similarity at the valley is low (Pascual-Marqui and Michel, 1995; Thomas et al., 2002; Walther, 2005), which means that the transition from one topographic map to another is completed at the negative peak of GFP.

- (2) The modified k-means clustering algorithm is used for cluster analysis (Murray et al., 2008), and EEG data is clustered into n microstates. Clustering model is as follows:

$$x_n = Az_n + \varepsilon_n \quad (13)$$

where, x_n is the EEG signal sampled for the n th time, $1 \leq n \leq N$, N is the number of time sample; $A \in R^{C \times K}$ is the topographic map of clustering, C is the number of channels, and K is the number of clustering (the number of microstate class). $z_n \in R^{K \times N}$ is the activation state of the microstate at the n th sampling; ε_n is EEG signal noise sampled at the n th time.

- (3) Global Explained Variance (GEV) and Cross-Validation criterion (Poulsen et al., 2020; CV) are calculated to evaluate the fitness of microstates and determine the optimal number of microstates.

GEV is an index to measure the similarity between each EEG sample and its assigned microstate, so the higher the GEV value, the better the result. The formula is as follows:

$$GEV_n = \frac{(\text{corr}(x_n, a_l)) \cdot GFP_n}{\sum_{n=1}^N GFP_n^2} \quad (14)$$

where, GFPn is the global field potential and the standard deviation of all EEG electrodes sampled at the nth time.

The value of CV is related to residual noise, so a smaller value of CV should be obtained. The calculating formula is as follows:

$$CV = \hat{\sigma}^2 \cdot \left(\frac{C-1}{C-K-1} \right)^2 \quad (15)$$

$$\hat{\sigma}^2 = \frac{\sum_{n=1}^N x_n^T x_n - (a_n^T x_n)^2}{N(C-1)} \quad (16)$$

where, $\hat{\sigma}$ is the estimator of the residual noise variance.

(4) After matching the extracted microstates to the EEG signals of the subjects, the EEG microstate parameters between the two tasks at each frequency band are calculated respectively:

- ① Mean Duration (MD): the mean duration of time that one microstate keeps stable.
- ② Time Coverage Ratio (TCR): percentage in time coverage of one type of microstate.
- ③ Occurrence Per Second (OPS): frequency of occurrence of one microstate.
- ④ Global Explained Variance (GEV): an index to measure the similarity between each EEG sample and its assigned microstate.

2.2. Data and pre-processing

The dataset used in this paper is derived from a novel cognition-guided neurofeedback BCI dataset on nicotine addiction, which includes smoking subjects performing two cognitively guided tasks at a sampling frequency of 250 Hz (Bu et al., 2021). The cognitively guided task of the dataset is to record EEG data by allowing subjects to focus on the smoking-related pictures (e.g., holding a cigarette in hand) and paired neutral pictures (e.g., holding a pencil in hand). The EEG data of smoking-related pictures and neutral pictures on each subject were recorded in six groups, respectively.

In this study, EEG data of 20 subjects were selected from this dataset, including 120 groups of EEG data in smoking-related pictures (smoke) and 120 groups of EEG data in neutral pictures (neutral). In the pre-processing step, the original EEG signals were filtered to 0.1–40 Hz to remove noise and other interference signals. At the same time, the eye electrodes were removed, and the corresponding potentials of 45 electrodes (F7, F5, F3, F1, FZ, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCZ, FC2, FC4, FC6, FT8, T7, C5, C3, C1, CZ, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPZ, CP2, CP4, CP6, TP8, P7, P5, P3, P1, PZ, P2, P4, P6, P8) were selected for evaluation.

By using the CEEMDAN and instantaneous frequency model in the improved HHT method, our experiment divided the addiction EEG data with two different tasks into five frequency bands including delta band (0.5–4 Hz), theta band (4–8 Hz), alpha band (8–12 Hz), beta band (12–30 Hz), and gamma band (30–40 Hz). Then, we analyzed and compared the microstate parameters at each frequency band.

3. Results

3.1. Full band microstates

According to GEV and CV, the difference is the largest when the number of EEG microstates in neutral group is 6 and smoke group is 5 in Figure 1. Therefore, we selected 6 microstates in neutral group and 5 in smoke group, which is shown in Figure 2. Microstates A, B, and C in neutral group and smoke group are similar to the classic microstates. However, microstate D, with positive and negative voltage located in the frontal central region (ignoring polarity), is related to attention network and sleep (Delorme and Makeig, 2004), which is split into microstates D1 and D2 in neutral group, but does not split in smoke group. In addition, an additional microstate E is generated in both groups.

3.2. Time-frequency analysis

In this study, as mentioned above, instantaneous parameters were extracted to provide an insight into the local variations in the spectral domain of EEG data. For each subject, and for each

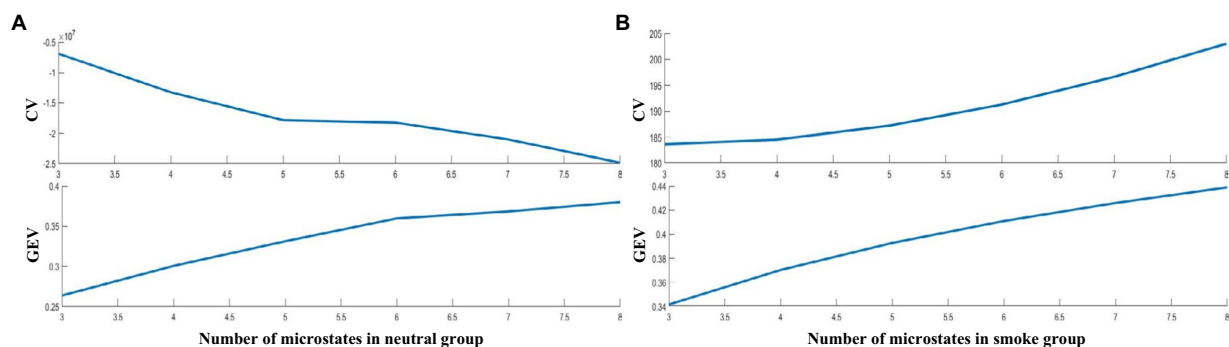


FIGURE 1
The effective number of microstates based on the fitness.

channel, the EEG data were decomposed into a set of IMFs using the CEEMDAN algorithm, which is followed by the estimation of instantaneous amplitudes and instantaneous frequency using HT. As an example, Figure 3 shows the decomposed IMFs for channel F7 of EEG data from a representative subject. Figure 4 shows the corresponding energies at each band. It should be noted that the whole-time length of 1 min is used for decomposition and for a better display only 5 s data are shown. Figure 5 shows the sub-band energies across 45 electrodes for one time instance.

3.3. Sub-band microstates and statistical analysis

We performed microstate analysis at each EEG frequency band obtained by improved HHT method. According to GEV and CV, the optimal number of microstate at delta, alpha and theta bands is 5, while the optimal number of microstate at beta and gamma bands is 4. Figure 6 shows each band microstate topographic maps obtained by improved HHT method.

Microstate parameters at each frequency band under the improved HHT method were calculated, including MD, OPS, TCR,

and GEV. The results are shown in Table 1. Multi-way ANOVA was performed for microstate parameters at each frequency band. We find significant class \times group interactions for all microstate parameters: ① delta band: MD ($F=120.98$, $p<0.001$, $\eta^2=347164.93$), OPS ($F=702.13$, $p<0.001$, $\eta^2=140.94$), TCR ($F=551.46$, $p<0.001$, $\eta^2=5.55$), GEV ($F=305.58$, $p<0.001$, $\eta^2=4.78$); ② alpha band: GEV ($F=9.98$, $p<0.001$, $\eta^2=0.047$); ③ beta band: TCR ($F=3.42$, $p=0.017$, $\eta^2=0.04$), GEV ($F=25.45$, $p<0.001$, $\eta^2=0.16$); Then, we performed separate one-way ANOVA for each microstate parameter at delta, alpha and beta bands between neutral and smoke group. The results are shown in “Supplementary Tables A.” These follow-up tests reveals significant between group differences for band microstates: ① At delta band, the OPS, TCR and GEV of microstate D5 in neutral group are higher than those in smoke group ($OPS_{\text{neutral}}=1.59\pm0.30$, $OPS_{\text{smoke}}=1.44\pm0.20$; $TCR_{\text{neutral}}=0.22\pm0.11$, $TCR_{\text{smoke}}=0.18\pm0.03$; $GEV_{\text{neutral}}=7.75\pm2.49$, $GEV_{\text{smoke}}=6.64\pm1.81$); ② At alpha band, the OPS, TCR and GEV of microstate A2 in neutral group are lower than those in smoke group ($OPS_{\text{neutral}}=2.03\pm0.22$, $OPS_{\text{smoke}}=2.14\pm0.15$; $TCR_{\text{neutral}}=0.21\pm0.03$, $TCR_{\text{smoke}}=0.22\pm0.02$; $GEV_{\text{neutral}}=7.62\pm1.65$, $GEV_{\text{smoke}}=8.62\pm1.65$); ③ At beta band, the TCR and GEV of microstate B2 in neutral group are lower than those in smoke group ($TCR_{\text{neutral}}=0.25\pm0.04$, $TCR_{\text{smoke}}=0.27\pm0.02$; $GEV_{\text{neutral}}=7.47\pm1.89$, $GEV_{\text{smoke}}=8.67\pm3.53$).

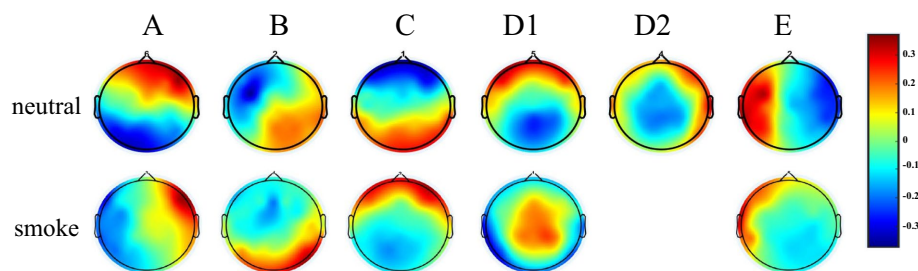


FIGURE 2
EEG microstates on two response tasks, the number in neutral group is 6, the number in smoke group is 5.

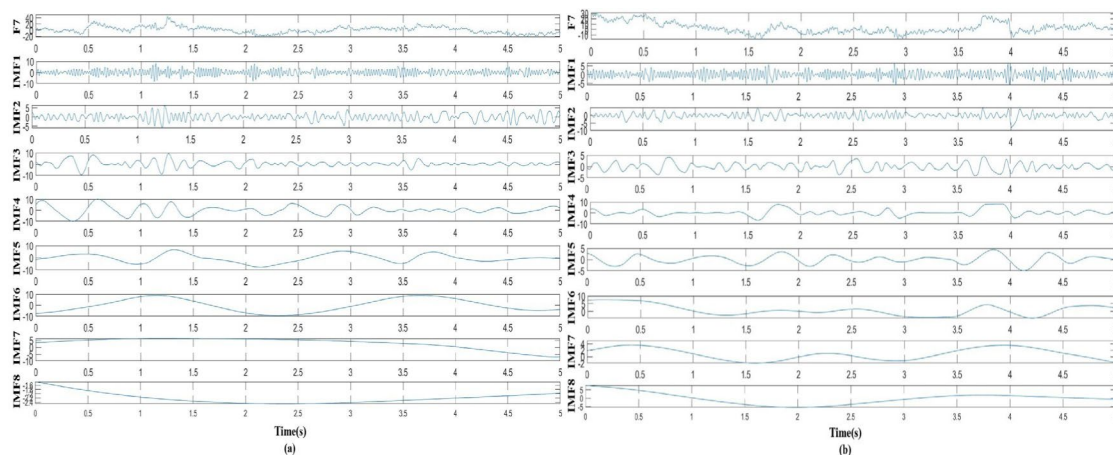


FIGURE 3
Taking channel F7 as an example, the EEG signal was decomposed into IMFs, where (A) decomposed signal in neutral group, (B) decomposed signal in smoke group.

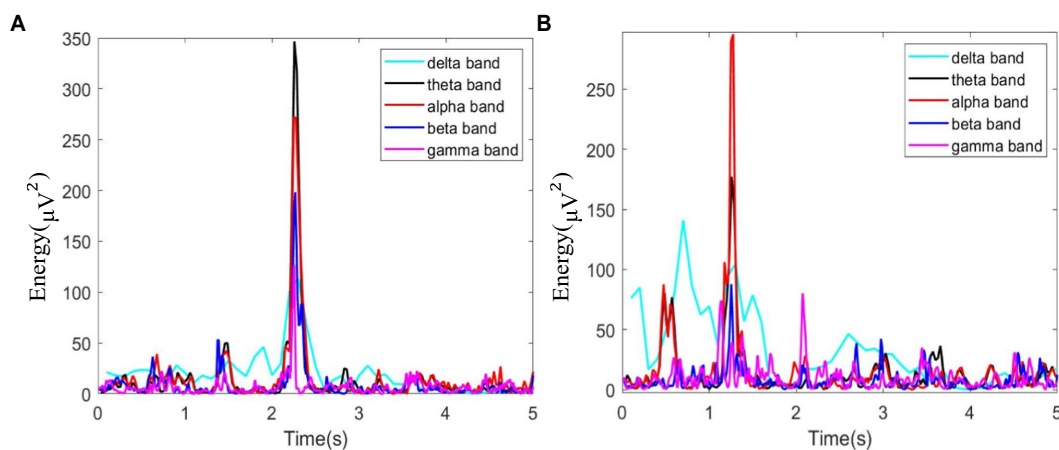


FIGURE 4
Energy diagram at each band of F7 channel within 5s, where (A) neutral group; (B) smoke group.

3.4. Classification and recognition on microstate parameters

According to the results in section 3.3, microstate parameters with significant differences between the two tasks were selected as features, including MD, OPS, TCR and GEV at delta, alpha, and beta band, which were performed for classification under Gaussian kernel SVM classifier.

The results are shown in Table 2. The classification effect is optimal at delta band. The microstate D1 has the highest classification accuracy (92%), sensitivity (94%) and specificity (91%). Other microstates at delta band also have better classification results. Then, microstate A2 and A3 at alpha band also have good classification effect, with the highest accuracy of 78%, sensitivity of 75% and specificity of 87%. Furthermore, microstate B1 and B2 at beta band have general classification effect, with accuracy of 73%, sensitivity of 90% and specificity of 84%.

Besides, selecting the microstates D1, A2, and B2 with the best classification result, we also plot the ROC curve, which shows that D1 has the best results in classification. The specific result is shown in Figure 7.

3.5. Comparing with other methods in nicotine addiction detection

Some EEG analysis methods which similar to the improved HHT-Microstate were also employed to analyze EEG data of these nicotine addiction subjects. Previous studies have greatly improved the decoding accuracy of EEG by calculating the spatial feature of the Riemann distance in the EEG of motion imagination at frequency bands (Qu et al., 2022). And in the traditional microstate analysis, the Finite Impulse Response (FIR) filter in EEGLAB (Michel and Koenig, 2017) was employed to filter EEG data according to the frequency band range. Therefore, microstates and Riemann distance were calculated from EEG signal at each frequency band filtering by FIR and HHT. The result of analysis and comparison according to these methods are as follows.

Figure 8 shows each band microstate topographic maps obtained by FIR method. For further comparison with the improved HHT-Microstate, similarity index (Kingsley and Sethukarasi, 2023) was calculated for each single-band and full-band microstate topographic maps, respectively, for the purpose of corresponding comparison.

Firstly, the difference index of each EEG band topographic maps between the two types of tasks was compared. The results in Figure 9 show that the improved HHT method provides more variability among the topographic maps at each frequency band.

Secondly, the permutation test was conducted for the similarity of the two topographic maps (Koenig et al., 1999). Table 3 shows the permutation test results of the similarity index among topographic maps. The difference of the test results under the improved HHT method is mainly reflected in the microstate A2 and A5 at alpha band and the microstate B4 at beta band.

In addition, the GEV under the two methods were calculated, respectively. The GEV under the improved HHT method is higher than the traditional filtering method, which are shown in Table 4.

Furthermore, microstate parameters and Riemann distance at each frequency band calculated under the FIR method were shown in “Supplementary Tables B,” which include MD, OPS, TCR and GEV for microstate, model of AIRM, Stein, Jeffery and LogED for Riemann distance. Multi-way ANOVA was performed for each parameter at each frequency band. We do not find significant class \times group interactions for all parameters. We also did one-way ANOVA for each microstate parameter and Riemann distance at delta, alpha and beta bands between neutral and smoke group. The results were shown in “Supplementary Tables C.” Only few parameters have significant difference.

Finally, we also select microstate parameters and Riemann distance at delta, alpha and beta bands between the two tasks as features, which were performed for classification under Gaussian kernel SVM classifier.

The results are shown in Tables 5–7. The optimal effect of classification for microstate is A4 at alpha band, which has the highest classification accuracy (87%), sensitivity (88%) and specificity (86%), and for Riemann distance it is beta band, which has the highest classification accuracy (71%), sensitivity (67%) and specificity (75%).

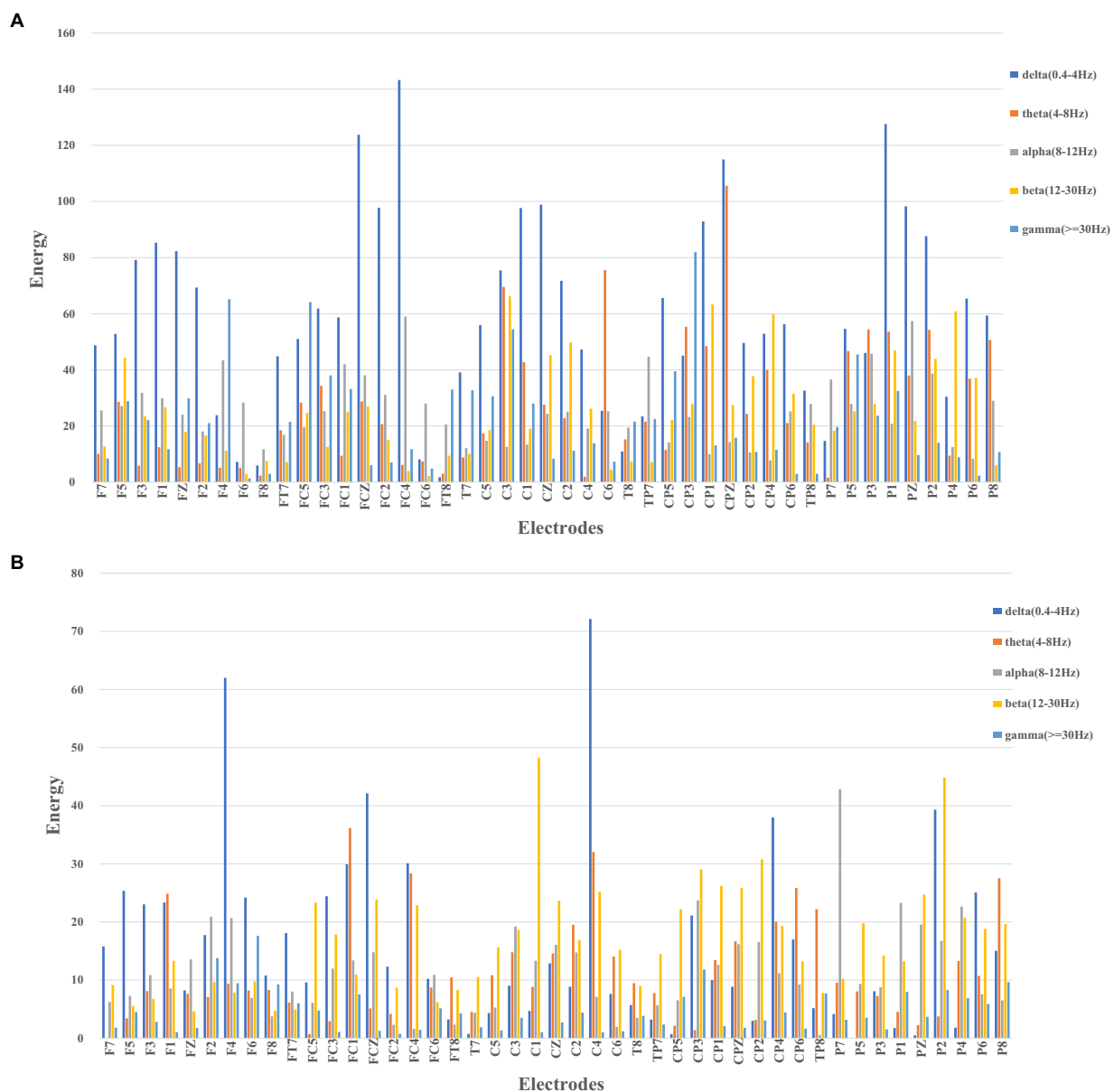


FIGURE 5
Sub-band energies across 45 electrodes, where (A) neutral group; (B) smoke group.

The effect of classification for FIR-Microstate and FIR-Riemann is inferior to improved HHT-Microstate.

4. Discussion

As a category of mental illness, substance addiction is a cause of avoidable morbidity and mortality around the world. Nicotine addiction is the most widely distributed and the most numerous substance addiction type. According to relevant studies, nicotine addicts are different from non-addicts in cognitive function, sleep structure and smoking cue response. Therefore, many studies mainly carry out on the mechanism of addiction and intervention methods, which have great potential clinical benefits for the intervention and treatment of substance addiction. Spontaneous

EEG signal, modulated by cognitive and sensory processing (Samaha et al., 2022), fluctuates in milliseconds and explains the transient brain functional states. Therefore, it is necessary to further explore the brain mechanism of smoking cue response and find effective markers of smoking cue response as targets for addiction detection and intervention.

In order to determine the fluctuation dynamics of brain neural sources in the time domain, Lehman (Lehmann et al., 1987; Lehmann, 1994) proposed the method of EEG microstate analysis which quantified the spatial distribution of nerve potentials among scalp electrodes at each time, reflected the sum of instantaneous activity of brain neutral clusters with fewer microstate topographic maps, and examined the functional activity network of the brain (Lehmann et al., 1987). It is the best choice for time domain analysis. On this basis, Ehtasham et al. (2019) used the HHT

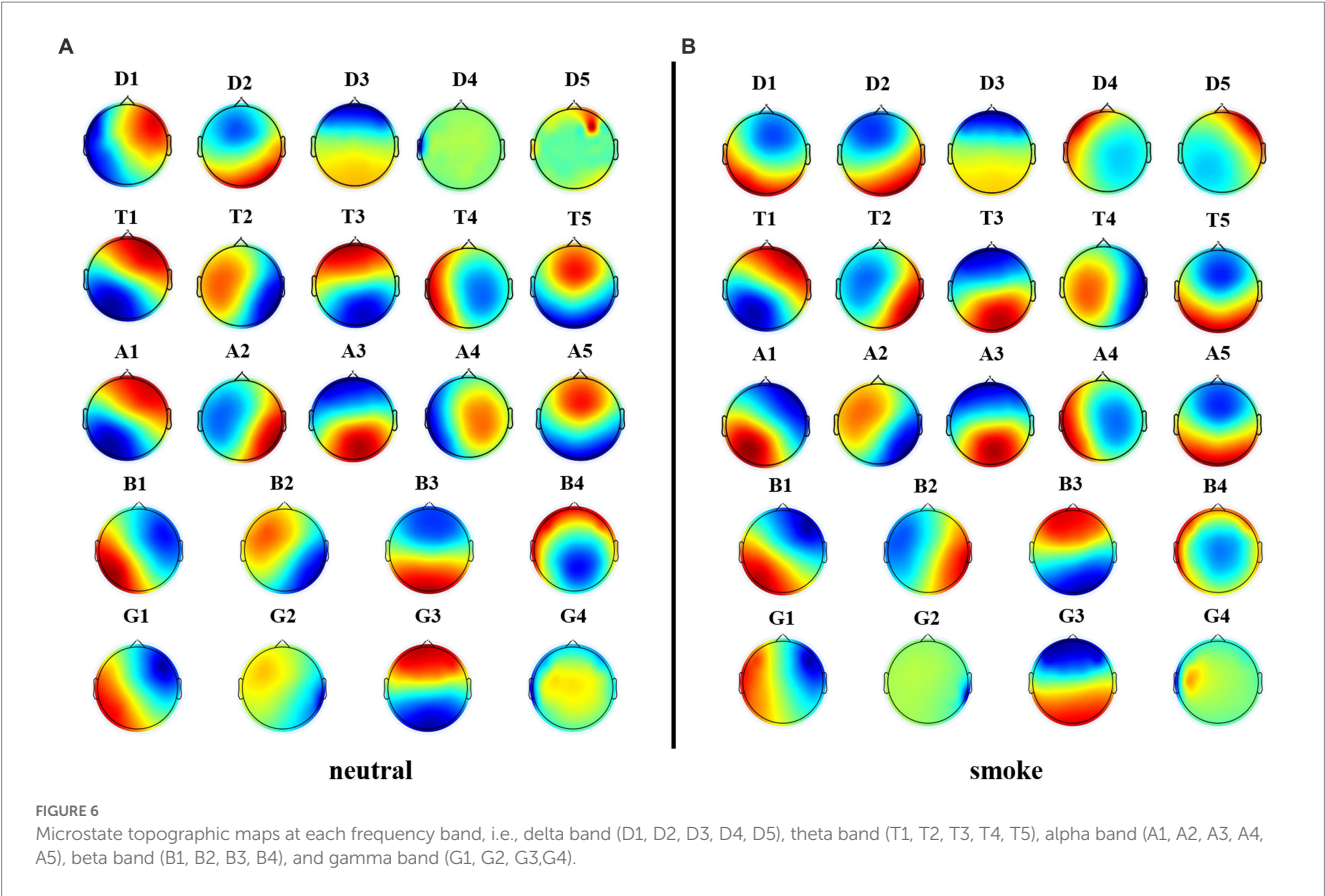
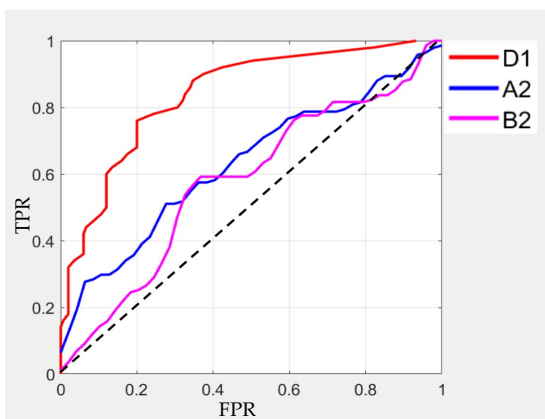


TABLE 1 Microstate parameters at each frequency band under the improved HHT method.

Subjects	Microstates	MD		OPS		TCR		GEV	
		neutral	smoke	neutral	smoke	neutral	smoke	neutral	smoke
Delta	D1	127.66 ± 10.97	131.01 ± 14.50	1.62 ± 0.39	1.63 ± 0.40	0.21 ± 0.06	0.22 ± 0.07	19.30 ± 6.60	20.33 ± 9.49
	D2	125.50 ± 7.63	127.18 ± 7.83	1.49 ± 0.49	1.68 ± 0.20	0.19 ± 0.07	0.21 ± 0.03	11.77 ± 2.33	12.54 ± 3.79
	D3	123.63 ± 9.19	126.21 ± 6.37	1.45 ± 0.42	1.59 ± 0.16	0.18 ± 0.06	0.20 ± 0.03	9.55 ± 1.90	9.26 ± 1.93
	D4	127.82 ± 23.78	125.16 ± 7.34	1.57 ± 0.28	1.53 ± 0.20	0.21 ± 0.09	0.19 ± 0.03	9.19 ± 3.66	8.27 ± 1.96
	D5	130.38 ± 34.43	121.22 ± 5.64	1.59 ± 0.30	1.44 ± 0.20	0.22 ± 0.11	0.18 ± 0.03	7.75 ± 2.49	6.64 ± 1.81
Theta	T1	123.31 ± 7.28	122.23 ± 8.37	1.95 ± 0.17	1.89 ± 0.16	0.24 ± 0.03	0.23 ± 0.03	10.00 ± 1.93	10.04 ± 2.32
	T2	117.21 ± 5.98	119.86 ± 6.50	1.85 ± 0.11	1.82 ± 0.14	0.22 ± 0.02	0.22 ± 0.02	8.41 ± 1.44	8.45 ± 1.61
	T3	115.51 ± 4.88	118.03 ± 6.38	1.74 ± 0.12	1.74 ± 0.11	0.20 ± 0.02	0.21 ± 0.02	7.37 ± 1.37	7.59 ± 1.19
	T4	110.36 ± 5.43	112.61 ± 5.81	1.59 ± 0.13	1.64 ± 0.13	0.18 ± 0.02	0.18 ± 0.02	6.23 ± 0.96	6.40 ± 0.79
	T5	109.74 ± 5.89	106.56 ± 4.40	1.49 ± 0.17	1.51 ± 0.15	0.16 ± 0.02	0.16 ± 0.02	5.29 ± 0.84	5.27 ± 0.89
Alpha	A1	108.60 ± 6.93	105.67 ± 5.46	2.20 ± 0.20	2.21 ± 0.17	0.24 ± 0.03	0.23 ± 0.02	9.64 ± 2.22	9.56 ± 1.50
	A2	101.13 ± 6.60	104.07 ± 3.83	2.03 ± 0.22	2.14 ± 0.15	0.21 ± 0.03	0.22 ± 0.02	7.62 ± 1.65	8.62 ± 1.65
	A3	100.27 ± 6.50	102.84 ± 5.59	1.96 ± 0.19	2.01 ± 0.17	0.20 ± 0.03	0.21 ± 0.02	7.18 ± 1.90	7.41 ± 1.47
	A4	97.52 ± 4.45	97.24 ± 6.64	1.85 ± 0.18	1.79 ± 0.15	0.18 ± 0.02	0.17 ± 0.22	6.13 ± 0.82	5.97 ± 1.03
	A5	96.73 ± 5.82	95.94 ± 4.41	1.81 ± 0.22	1.70 ± 0.15	0.18 ± 0.03	0.16 ± 0.02	5.69 ± 1.69	5.16 ± 0.90
Beta	B1	87.81 ± 6.92	83.90 ± 16.97	3.39 ± 0.16	3.88 ± 2.70	0.30 ± 0.03	0.29 ± 0.03	9.68 ± 3.47	9.48 ± 3.55
	B2	80.62 ± 5.13	80.14 ± 14.76	3.11 ± 0.45	3.81 ± 2.58	0.25 ± 0.04	0.27 ± 0.02	7.47 ± 1.89	8.67 ± 3.53
	B3	80.50 ± 3.65	76.48 ± 14.88	1.62 ± 0.39	1.63 ± 0.40	0.25 ± 0.03	0.23 ± 0.03	6.60 ± 1.17	6.35 ± 1.39
	B4	76.90 ± 5.42	76.24 ± 15.44	1.49 ± 0.49	1.68 ± 0.20	0.20 ± 0.03	0.21 ± 0.03	4.91 ± 1.18	5.25 ± 1.33
Gamma	G1	120.30 ± 52.82	138.77 ± 102.42	1.45 ± 0.42	1.59 ± 0.16	0.30 ± 0.07	0.31 ± 0.08	10.09 ± 5.82	10.83 ± 7.82
	G2	106.48 ± 53.52	93.68 ± 8.23	1.57 ± 0.28	1.53 ± 0.20	0.27 ± 0.05	0.26 ± 0.04	7.18 ± 2.88	6.52 ± 1.84
	G3	90.54 ± 12.54	87.53 ± 20.00	1.59 ± 0.30	1.44 ± 0.20	0.22 ± 0.05	0.22 ± 0.08	4.68 ± 1.28	4.71 ± 2.10
	G4	101.80 ± 57.41	94.04 ± 26.97	1.95 ± 0.17	1.89 ± 0.16	0.21 ± 0.05	0.21 ± 0.06	3.98 ± 1.17	4.51 ± 1.64

TABLE 2 Classification results of microstates at delta, alpha, and beta bands by using improved HHT-Microstate.

Subjects	Microstates	Accuracy (%)	Sensitivity (%)	Specificity (%)
Delta [0.1–4 Hz]	D1	92.86	94.29	91.43
	D2	90.35	90.03	90.29
	D3	83.33	68.89	91.11
	D4	87.88	86.02	88.14
	D5	83.34	66.67	91.36
Alpha [8–12 Hz]	A1	69.44	75.08	63.89
	A2	78.30	73.58	83.02
	A3	71.28	46.81	87.74
	A4	56.71	64.86	48.16
	A5	61.86	58.28	64.22
Beta [12–30 Hz]	B1	70.19	55.77	84.62
	B2	73.47	90.88	53.06
	B3	61.02	61.02	61.02
	B4	63.64	70.45	56.82

FIGURE 7
ROC curves of classification on microstate D1, A2, and B2..

method to transform the time-domain data into the spectral domain data, which retained its instantaneous characteristics, constructed the correlation between microstate and spectral features, and proved that this method was superior to the traditional filtering method through experimental comparison. Therefore, in this paper, an improved HHT method and FIR method were used, respectively, for frequency band decomposition and microstate analysis on EEG of nicotine addiction under two types of different tasks. Besides, microstate parameters with significant differences after improved HHT decomposition were used as features to classify and detect nicotine addiction.

4.1. Comparison of microstate parameters at each frequency band

Since cue response plays an important role in the psychological cognition, withdrawal and relapse of nicotine

addictions, many experiments have been conducted to study the different cues provided by nicotine addicts, which include the cues related to neutral control and cigarettes to different degrees. Through different cue feedback, it is found that cigarette-related cues caused higher levels of de-alpha synchronization (Cui et al., 2013), and theta band in frontal lobe shows strong network coherence in smoking cues (Shinan, 2020). At the same time, studies have shown that there are significant differences in microstate parameters between addicts and non-addicts, and better identification and detection can be achieved under the SVM optimized by genetic algorithm (Peng, 2019). Thus, combined with previous studies, whether there are significant differences in microstate parameters at different frequency bands and whether accurate identification and detection can be achieved in the face of different smoking cues is a worth studying problem in the cue response of substance addicts.

Therefore, in our paper, two time-frequency decomposition methods were employed to divide the EEG data of nicotine addicts into five different bands and compared the microstate parameters between two different cue-response tasks at each frequency band to find differences.

The experimental results show that the two groups of full band microstates in Figure 2, microstate D in neutral group split into microstates D1 and D2, while those in smoke group does not split. In addition, an additional microstate E is generated in both tasks groups. Whether this microstate is unique to nicotine addicts needs further research. Then, MD, OPS, TCR and GEV at each frequency band obtained by the improved HHT method are shown in Table 1. According to the result of multi-way ANOVA, we find significant class \times group interactions for microstate D5, A2 and B2 at specific delta, alpha and beta bands. However, the study of Ehtasham et al. (2019) found that the optimal number of microstates at each frequency band was 4, and the parameters were similar and consistent at each frequency band, without significant difference through healthy non-addicts. Our experimental results are similar to previous studies that there are significant differences in coherence, power, and energy between addicts and non-addicts at specific EEG bands (Reid et al.,

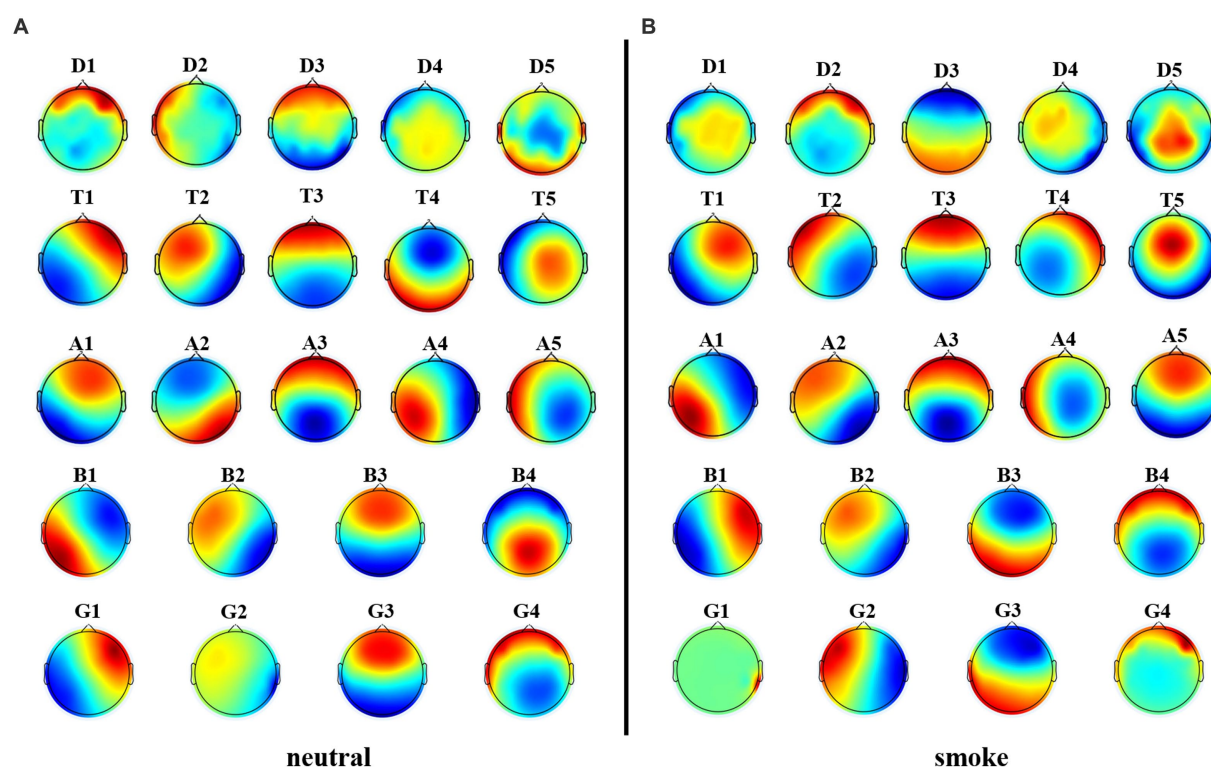


FIGURE 8

Microstate topographic maps at each frequency band, i.e., delta band (D1, D2, D3, D4, D5), theta band (T1, T2, T3, T4, T5), alpha band (A1, A2, A3, A4, A5), beta band (B1, B2, B3, B4), and gamma band (G1, G2, G3, G4).

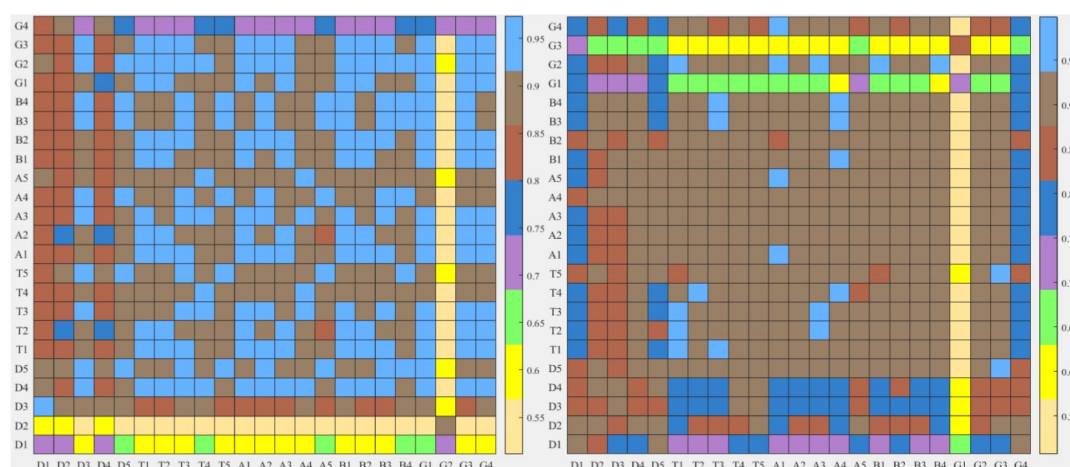


FIGURE 9

After frequency band division by the two methods, the similarity index hot plot of all microstate topographic maps at each EEG frequency band between two kinds of different tasks is extracted. The horizontal axis is EEG microstate of the neutral group and the vertical axis is EEG microstate of the smoke group.

2004; Peng, 2019), which may indirectly indicate that the EEG of nicotine addicts is different at certain frequency bands. Therefore, the difference of EEG microstate parameters at specific frequency bands can be used to detect substance addiction. Furthermore, based on the results of one-way ANOVA for each microstate parameter between neutral and smoke group at delta, alpha and beta bands, $D5_{OPS}$, $D5_{TCR}$

and $D5_{GEV}$ in neutral group at delta band are higher than those in smoke group, $A2_{OPS}$, $A2_{TCR}$ and $A2_{GEV}$ in neutral group at alpha band are lower than those in smoke group and $B2_{TCR}$, $B2_{GEV}$ in neutral group at beta band are higher than those in smoke group. We can distinguish and detect nicotine addiction with different cue responses mainly by these microstate parameters at these frequency bands.

TABLE 3 The permutation test results of the similarity index between the full-band and single-band microstate topographic map of subjects' EEG.

Microstates		A		B		C		D		E	
Bands		Improved HHT	FIR	Improved HHT	FIR	Improved HHT	FIR	Improved HHT	FIR	Improved HHT	FIR
Delta	D1	0.31	0.85	0.28	0.86	0.34	0.88*	0.29	0.83	0.35	0.86*
	D2	0.25	0.86	0.22	0.86	0.27	0.88	0.23	0.83	0.34	0.85
	D3	0.23	0.86	0.20	0.86	0.24	0.88	0.22	0.84	0.32	0.86
	D4	0.30	0.87	0.38	0.87	0.39	0.89	0.38	0.84	0.35	0.87
	D5	0.37	0.87	0.35	0.87	0.32	0.89	0.37	0.84	0.38	0.87
Theta	T1	0.25	0.79	0.23*	0.79	0.27	0.85	0.23*	0.76	0.34	0.92
	T2	0.13	0.80	0.13	0.80	0.16	0.86	0.16	0.77	0.22	0.92
	T3	0.28	0.79	0.29	0.79	0.26	0.85	0.30	0.76	0.22	0.92
	T4	0.30	0.81	0.24	0.80	0.39	0.86	0.25	0.77	0.42	0.91
	T5	0.20	0.82	0.15	0.81	0.24	0.87	0.19	0.78	0.29	0.92
Alpha	A1	0.47	0.80	0.47	0.79	0.47	0.85	0.48	0.77	0.33	0.92
	A2	0.04*	0.80	0.04	0.79	0.04*	0.85	0.05*	0.76	0.04*	0.90
	A3	0.30	0.81	0.34	0.80	0.37	0.86	0.32	0.77	0.35	0.92
	A4	0.29	0.80	0.30	0.79	0.29	0.85	0.32	0.77	0.28	0.91
	A5	0.21*	0.80	0.19*	0.80	0.17*	0.86	0.20*	0.77	0.11*	0.91
Beta	B1	0.18	0.79	0.19	0.80	0.20	0.83	0.14	0.77	---	---
	B2	0.17	0.79	0.17	0.80	0.16	0.82	0.16	0.77	---	---
	B3	0.32	0.80	0.26	0.80	0.23	0.85	0.24	0.77	---	---
	B4	0.50*	0.81	0.04*	0.81	0.06*	0.85	0.04*	0.78	---	---
Gamma	G1	0.16	0.76	0.25	0.79	0.21	0.78	0.21*	0.76	---	---
	G2	0.22	0.76	0.34	0.80	0.30	0.78	0.28	0.77	---	---
	G3	0.32	0.76	0.31	0.79	0.21	0.78	0.11	0.76	---	---
	G4	0.40	0.80	0.37	0.81	0.27	0.83	0.16	0.78	---	---

The results with statistically significant differences are indicated by asterisks (*).

TABLE 4 GEV for all microstates at each frequency band under improved HHT method and traditional filtering method.

EEG Data	Methods	Delta	Theta	Alpha	Beta	Gamma
Neutral	Improved HHT	55.17	37.93	36.77	27.10	21.15
	FIR	31.79	33.20	34.89	24.50	33.19
Smoke	Improved HHT	55.14	38.29	37.15	27.57	22.87
	FIR	32.49	32.8	34.99	28.39	34.21

4.2. Detection of nicotine addiction by improved HHT-microstate method

Previous studies have used microstate correlation parameters to classify and detect different diseases or tasks and have achieved better identification and detection effect. For example, for heroin addicts, the features of EEG microstate parameters and negative peak of microstate duration were used to classify, with the accuracy of 72% (Peng, 2019). At the same time, there are different spatial microstates between patients with high and normal cranial pressure. Microstate parameters were used to classify patients with high and low cranial pressure, which can obtain the highest classification accuracy (87%; Shuaiyang, 2021). In this paper, microstate parameters with different frequency bands under

different cue-response tasks were selected as features to the Gaussian kernel SVM classifier for classification and detection. The microstate parameters at delta band, alpha band and beta band were used to classify substance addiction. The microstate D1 at delta band has the highest classification accuracy (92%), sensitivity (94%), and specificity (91%), the microstate A2 at alpha band and microstate B2 at beta band also have better classification result. At the same time, the microstates with the best classification result at each band were selected and plot the ROC curves, which also mainly reflected the best result on microstate D1 and the better result on microstate A2 and B2 by evaluating the AUC under each curve. Therefore, microstate parameters at delta and alpha bands are promising for the identification and detection of nicotine addiction.

TABLE 5 Classification results of microstates at delta, alpha, and beta bands by using FIR-Microstate.

Subjects	Microstates	Accuracy (%)	Sensitivity (%)	Specificity (%)
Delta [0.1–4 Hz]	D1	64.44	62.22	66.67
	D2	66.67	68.89	64.44
	D3	62.22	93.33	31.11
	D4	77.78	86.67	31.11
	D5	77.78	73.33	82.22
Alpha [8–12 Hz]	A1	54.00	60.00	48.00
	A2	65.00	52.00	78.00
	A3	63.00	62.00	64.00
	A4	87.00	88.00	86.00
	A5	79.00	64.00	94.00
Beta [12–30 Hz]	B1	65.91	81.82	50.00
	B2	64.78	72.73	56.82
	B3	64.77	72.73	56.82
	B4	65.91	77.27	54.55

TABLE 6 Classification results of Riemann distance at delta, alpha, and beta bands by using HHT-Riemann.

Bands	Accuracy (%)	Sensitivity (%)	Specificity (%)
Delta	62.22	55.56	68.89
Alpha	57.78	64.44	51.11
Beta	64.29	71.43	57.14

TABLE 7 Classification results of Riemann distance at delta, alpha, and beta bands by using FIR-Riemann.

Bands	Accuracy (%)	Sensitivity (%)	Specificity (%)
Delta	60.00	51.11	68.89
Alpha	57.45	57.45	57.45
Beta	71.43	67.35	75.51

4.3. Comparison of analysis result between the improved HHT-microstate and other methods

In order to further prove the intrinsic superiority of the improved HHT-Microstate method, this experiment compared the improved HHT-Microstate method with other similar EEG analysis methods, including the frequency band microstates and frequency band Riemann distance extracted by FIR filtering methods, and conducted corresponding statistical analysis and classification detection, respectively. According to the comparison results in section 3.5, it is found that there were significant differences on the similarity of microstate topographic maps, statistical analysis and classification results of parameters between different methods.

Firstly, according to Figures 6, 8, it could obviously observe that there were significant differences between the two methods for each EEG microstate topographic map at delta and gamma bands in neutral and smoke group, and there were also some other significant differences in the microstate topographic maps at other bands. The similarity index among all topographic maps at each EEG frequency band of the two groups obtained by the improved HHT and FIR is shown in Figure 9, in which the improved HHT method has a lot of variability among the band topographic maps. According to the permutation test in Table 3, the similarity of improved HHT-Microstate method at alpha and beta bands is significantly different. However, the FIR method does not detect these differences. In addition, the GEV under the improved HHT-Microstate method is higher than the FIR method.

Then, the same multi-way ANOVA as improved HHT-Microstate method was performed on the FIR band microstate parameters and Riemann distance, however, there were no significant interaction. At the same time, only a few parameters of each feature were significantly different under the one-way ANOVA between neutral and smoke group. For more accurate verification, we also chose feature parameters consistent with improved HHT-Microstate method for classification detection, which means that microstate parameters and Riemann distance at delta, alpha and beta band were selected as features for classification. According to the results in Tables 5–7, it is found that the result of classification for microstate is A4 at alpha band, which has the best accuracy (87%), sensitivity (88%) and specificity (86%), and for Riemann distance is beta band, which has the best accuracy (71%), sensitivity (67%) and specificity (75%). Thus, the effect of classification for the improved HHT-Microstate is better than FIR-Microstate, HHT-Riemann and FIR-Riemann methods, which means that the improved HHT-Microstate method is more suitable to represent the characteristics of EEG microstates and more representative than other methods in describing the dynamic characteristics of EEG.

4.4. Expectation

As a widespread medical and social problem in the world, substance addiction causes great harm to the physical health of

human and the stability of society. At present, the main treatment methods are physical therapy, drug therapy, psychological therapy and neurofeedback therapy. In recent years, with the development of BCI, EEG research has become a new diagnostic basis and treatment for substance addiction, which includes analyzing and comparing the differences of EEG signals in substance addiction, addiction withdrawal, and healthy controls. In this paper, the improved HHT method was used to divide the frequency band of EEG data and preserve the instantaneous characteristics of the time-domain data in the spectrum domain. At the same time, the EEG microstates of patients with nicotine addiction under different cue-response tasks were compared, and it was found that there were significant differences in the EEG microstates between different tasks.

However, with the advancement of computer technology, an increasing number of computational methods have been applied to brain research, such as Generative Adversarial Network (GAN), which solves the problem of imbalanced medical images (Hu et al., 2020), constructs super-resolution MR Images (You et al., 2022), reconstructs the lost BOLD signal (Yan et al., 2020) and fuse multi-modality medical images (Hu et al., 2021). All of them are latest research results in the brain science field and very enlightening the research of nicotine addiction in this paper. Therefore, our next step is to use these new techniques to analyze EEG signals and discover hidden information, which can be combined with microstate analysis.

Finally, there are individual differences in EEG signals. Increasing the amount of data will help further validate the results of this article, which is also one of our next steps. Besides, the methods of signal process could also be improved, for instance, some other adaptive time-frequency analysis methods (Wacker and Witte., 2011; Hadjileontiadis et al., 2017) can replace the improved HHT method and calculate the corresponding instantaneous frequency or instantaneous amplitude to obtain the unique EEG bands.

5. Conclusion

In this paper, we compared the difference of EEG microstates between nicotine addicts by using the improved HHT time-frequency decomposition method. We selected microstate parameters with significant difference as features for classification and got better recognition detection results. These results indicate that the EEG data at frequency bands obtained by the improved HHT method is more suitable to represent the characteristics of EEG signals, and the microstates obtained by this method can be effectively distinguished from the EEG data of nicotine addiction, which means that the improved HHT-Microstate analysis can offer new ideas and insights for the brain research of nicotine addiction and provide more effective methods and basis for the diagnosis and treatment of substance addiction.

References

Andreou, C., Faber, P. L., Leicht, G., Schoettle, D., Polomac, N., Hanganu-Opatz, I. L., et al. (2014). Resting-state connectivity in the prodromal phase of schizophrenia: insights from EEG microstates. *Schizophr. Res.* 152, 513–520. doi: 10.1016/j.schres.2013.12.008

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

XX: conceptualization and project administration. JF: methodology. YZ: software. DW: formal analysis. SY: investigation. CW: data curation. JF and RL: writing—original draft preparation. JH: supervision and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Nature Science Foundation of China, Yunnan Fundamental Research Projects and Scientific Research Fund of Hanshan Normal University, grant numbers 82060329, 202201AT070108, XY202106, and QD2021218.

Acknowledgments

Our research was thanked by the National Natural Science Foundation of China and Yunnan Fundamental Research Projects.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2023.1174399/full#supplementary-material>

Arjun, K., Alvaro, P. L., Christoph, M. M., and Faranak, F. (2014). Microstates in resting-state EEG: current status and future directions. *Neurosci. Biobehav. Rev.* 49, 105–113. doi: 10.1016/j.neubiorev.2014.12.010

- Arshad, J., Qaisar, A., Rehman, A.-U., Shakir, M., Nazir, M. K., Rehman, A. U., et al. (2022). Intel-ligent control of robotic arm using brain computer Interface and artificial intelligence. *Appl. Sci.* 12:10813. doi: 10.3390/app122110813
- Bao, C., Hong, H., Li, Z. X., and Zhu, X. (2009). Time-varying system identification using a newly improved HHT algorithm. *Comput. Struct.* 87, 1611–1623. doi: 10.1016/j.compstruc.2009.08.016
- Benos, J., and Kapinas, K. (1980). EEG examination in heroin addicts in rehabilitation. *Med. Welt* 31, 1395–1399.
- Bjork, J. M., and Gilman, J. M. (2014). The effects of acute alcohol administration on the human brain: insights from neuroimaging. *Neuropharmacology* 84, 101–110. doi: 10.1016/j.neuropharm.2013.07.039
- Britz, J., Van, D. V. D., and Michel, C. M. (2010). BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *Neuroimage* 52, 1162–1170. doi: 10.1016/j.neuroimage.2010.02.052
- Bu, J., Liu, C., Gou, H., Gan, H., and Zhang, X. (2021). A novel cognition-guided Neurofeedback BCI dataset on nicotine addiction. *Front. Neurosci.* 15:647844. doi: 10.3389/fnins.2021.647844
- Chu, C., Zhang, Z., Wang, J., Liu, S., Wang, F., Sun, Y., et al. (2021). Deep learning reveals personalized spatial spectral abnormalities of high delta and low alpha bands in EEG of patients with early Parkinson's disease. *J. Neural Eng.* 18:066036. doi: 10.1088/1741-2552/ac40a0
- Colrain, I. M., Nicholas, C. L., and Baker, F. C. (2014). Alcohol and the sleeping brain. *Handb. Clin. Neurol.* 125, 415–431. doi: 10.1016/B978-0-444-62619-6.00024-0
- Conroy, D. A., and Arnedt, J. T. (2014). Sleep and substance use disorders: an update current. *Psychiatry Rep.* 16, 487–494. doi: 10.1007/s11920-014-0487-3
- Coullaut-Valera, R., Arbaiza, I., Bajo, R., Arrúe, R., López, M. E., Coullaut-Valera, J., et al. (2014). Drug polyconsumption is associated with increased synchronization of brain electrical-activity at rest and in a counting task. *Int. J. Neural Syst.* 24:1450005. doi: 10.1142/S0129065714500051
- Cui, Y., Versace, F., Engelmann, J. M., Minnix, J. A., Robinson, J. D., Lam, C. Y., et al. (2013). Alpha oscillations in response to affective and cigarette-related stimuli in smokers. *Nicotine Tob. Res.* 15, 917–924. doi: 10.1093/ntr/nts209
- Daubechies, I., Lu, J., and Wu, H. T. (2011). Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.* 30, 243–261. doi: 10.1016/j.acha.2010.08.002
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Ehtasham, J., Pierpaolo, C., Filippo, Z., and Cosimo, D. G. (2019). Hilbert spectral analysis of EEG data reveals spectral dynamics associated with microstates. *J. Neurosci. Methods* 325:108317. doi: 10.1016/j.jneumeth.2019.108317
- Franken, I., Stam, C. J., Hendriks, V. M., and van den Brink, W. (2004). Electroencephalographic power and coherence analyses suggest altered brain function in abstinent male heroin-dependent patients. *Neuropsychobiology* 49, 105–110. doi: 10.1159/000076419
- Fulton, T. C., and Charlotte, A. B. (2009). Impulsivity, frontal lobes and risk for addiction. *Pharmacol. Biochem. Behav.* 93, 237–247. doi: 10.1016/j.pbb.2009.04.018
- Gabeff, V., Teijeiro, T., Zapater, M., Cammoun, L., Rheims, S., Rylvlin, P., et al. (2021). Interpreting deep learning models for epileptic seizure detection on EEG signals. *Artif. Intell. Med.* 117:102084. doi: 10.1016/j.artmed.2021.102084
- Gekht, A. B., Polunina, A. G., Briun, E. A., and Davydov, D. M. (2002). Brain bioelectrical activities in heroin addicts during early abstinence period. *Vserossiiskoe Obshchestvo Psikiatrov.* 442, 86–87. doi: 10.1007/s00428-002-0708-8
- Grieder, M., Koenig, T., Kinoshita, T., Utsunomiya, K., Wahlund, L. O., Dierks, T., et al. (2016). Discovering EEG resting state alterations of semantic dementia. *Clin. Neurophysiol.* 127, 2175–2181. doi: 10.1016/j.clinph.2016.01.025
- Hadjileontiadis, A., Leontios, J., Apostolidis, U., and Georgios, K. (2017). *Swarm decomposition: A novel signal analysis using swarm intelligence*. Signal Processing: The Official Publication of the EURASIP.
- Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., and Shen, Y. (2021). Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. *IEEE Trans. Med. Imaging* 41, 145–157. doi: 10.1109/TMI.2021.3107013
- Hu, S., Yu, W., Chen, Z., and Wang, S. (2020). Medical image reconstruction using generative adversarial network for Alzheimer disease assessment with class-imbalance problem. In: *IEEE International Conference on Computer and Communications*.
- Huang, N. E. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc London* 454, 903–995. doi: 10.21105/joss.02977
- Huang, Z. W., Long, S. R., Arnold, K. C., Chen, X., Blank, K., and Norden, E. (2009). On Instantaneous Frequency. *Adv. Adapt. Data Anal.* 01, 177–229. doi: 10.1142/S1793536909000096
- John, A. R., Cao, Z., Chen, H. T., Martens, K. E., Georgiades, M., Gilat, M., et al. (2023). Predicting the onset of freezing of gait using EEG dynamics. *Appl. Sci.* 13:302. doi: 10.3390/app13010302
- Kingsley, S., and Sethukarasi, T. (2023). Flower pollination student psychology optimization-integrated context deep learning and probabilistic-based fusion for image inpainting. *Int. J. Wavelets Multiresolut Inf. Process.* 21:3. doi: 10.1142/S0219691322500503
- Kiran, R. V., Rajagopalan, S. S., Bhardwaj, S., Panda, R., Reddam, V. R., Ganne, C., et al. (2018). Machine learning detects EEG microstate alterations in patients living with temporal lobe epilepsy. *Seizure* 61, 8–13. doi: 10.1016/j.seizure.2018.07.007
- Koenig, T., Lehmann, D., Merlo, M. C., Kochi, K., Hell, D., and Koukkou, M. (1999). A deviant EEG brain microstate in acute, neuroleptic-naïve schizophrenics at rest. *Eur. Arch. Psychiatry Clin. Neurosci.* 249, 205–211. doi: 10.1007/s004060050088
- Koenig, T., Studer, D., Hubl, D., Melie, L., and Strik, W. K. (2018). Brain connectivity at different time-scales measured with EEG. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1015–1024. doi: 10.1098/rstb.2005.1649
- Lehmann, D. (1994). Multichannel topography of human alpha EEG fields. *Electroencephalogr. Clin. Neurophysiol.* 31, 439–449. doi: 10.1016/0013-4694(71)90165-9
- Lehmann, D., Faber, P. L., and Galderisi, S. (2005). EEG microstate duration and syntax in acute, medication-naïve, first-episode schizo-phrenia: a multi-center study. *Psychiatry Res.* 138, 141–156. doi: 10.1016/j.psychres.2004.05.007
- Lehmann, D., Ozaki, H., and Pal, I. (1987). EEG alpha map series: rain microstates by space-oriented adaptive segmentation. *Electroencephalogr. Clin. Neurophysiol.* 67, 271–288. doi: 10.1016/0013-4694(87)90025-3
- Linyuan, D., and Xiaoyi, F. (2005). The neurobiological mechanism of nicotine dependence. *Adv. Psychol. Sci.* 13, 534–543. doi: 10.3969/j.issn.1671-3710.2005.04.018
- Littel, M., Franken, I., and Strien, J. V. (2009). Changes in the electroencephalographic Spectrum in response to smoking cues in smokers and ex-smoker. *Neuropsychobiology* 59, 43–50. doi: 10.1159/000205517
- Liu, S., Sun, Y., He, L., and Kang, Y. (2022). Weak signal processing methods based on improved HHT and filtering techniques for steel wire rope. *Appl. Sci.* 12:6969. doi: 10.3390/app12146969
- Michel, C. M., and Koenig, T. (2017). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. *Neuroimage* 180, 577–593. doi: 10.1016/j.neuroimage.2017.11.062
- Milz, P., Pascual-Marqui, R. D., Achermann, P., Kochi, K., and Faber, P. L. (2017). The EEG microstate topography is predominantly determined by intracortical sources in the alpha band. *Neuroimage* 162, 353–361. doi: 10.1016/j.neuroimage.2017.08.058
- Mumtaz, W., Vuong, P. L., Xia, L., Malik, A. S., and Rashid, R. B. A. (2017). An EEG-based machine learning method to screen alcohol use disorder. *Cogn. Neurodyn.* 11, 161–171. doi: 10.1007/s11571-016-9416-y
- Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr.* 20, 249–264. doi: 10.1007/s10548-008-0054-5
- Olivennes, A., Charlesnicolas, A., and Olievenstein, C. (1983). Changes in the waking electroencephalogram in serve heroin addiction. *Annales Medico-Psychologiques.* 4, 458–469.
- Pascual-Marqui, R. D., and Michel, C. M. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans. Biomed. Eng.* 42, 658–665. doi: 10.1109/10.391164
- Pascualmarqui, R. D., Michel, C. M., and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain international. journal of psychophysiology official. *J Int Organ Psychophysiol.* 18, 49–65. doi: 10.1016/0167-8760(84)90014-x
- Peng, R. (2019). *Microstate analysis and study of resting state EEG in heroin abusers*. Lanzhou: Lanzhou University.
- Pengfei, W., Ruiting, Y., Xin, M., and Hong, Z. (2019). Impulse or habit? The nature and mechanism of impulsivity in different stages of addiction. *Adv. Psychol. Sci.* 5, 834–842. doi: 10.3724/SP.J.1042.2019.00834
- Poulsen, A. T., Pedroni, A., Langer, N., and Hansen, L. K. (2020). Microstate EEGlab toolbox: An introductory guide. *bioRxiv* [Preprint].
- Prasanth, T., Thomas, J., Yuvaraj, R., Jing, J., Cash, S. S., Chaudhari, R., et al. (2020). Deep learning for Interictal Epileptiform spike detection from scalp EEG frequency sub bands. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020, 3703–3706. doi: 10.1109/EMBC44109.2020.9175644
- Qu, T., Jin, J., Xu, R., Wang, X., and Cichocki, A. (2022). Riemannian distance based channel selection and feature extraction combining discriminative time-frequency bands and Riemannian tangent space for MI-BCIs. *J. Neural Eng.* 19:056025. doi: 10.1088/1741-2552/ac9338
- Reid, M. S., Starosta, A., Flammio, F., Howard, B., and Pritchep, L. S. (2004). Quantitative electroencephalographic studies of cue-induced cocaine craving. *Clin. EEG Neurosci.* 34, 110–123. doi: 10.1177/155005940303400305
- Robbins, T. W., Ersche, K. D., and Everitt, B. J. (2008). Drug addiction and the memory systems of the brain. *Ann. N. Y. Acad. Sci.* 1141, 1–21. doi: 10.1196/annals.1441.020
- Samaha, J., Larocque, J. J., and Postle, B. R. (2022). Spontaneous alpha-band amplitude predicts subjective visibility but not discrimination accuracy during

- high-level perception. *Conscious. Cogn.* 102:103337. doi: 10.1016/j.concog.2022.103337
- Seitzman, B. A., Abell, M., Bartley, S. C., Erickson, M. A., and Hetrick, W. P. (2017). Cognitive manipulation of brain electric microstates. *Neuroimage* 146, 533–543. doi: 10.1016/j.neuroimage.2016.10.002
- Shinan, S. (2020). *Neural basis of smoking cue response and prediction of therapeutic effect of neurofeedback intervention*. Tianjin: Tianjin Normal University.
- Shuaiyang, L. (2021). *Study on intracranial pressure detection technique based on feature fusion of resting state EEG signals*. Zhengzhou: Zhengzhou University.
- Strik, W. K., Chiaramonti, R., Muscas, G. C., Paganini, M., Mueller, T. J., Fallgatter, A. J., et al. (1997). Decreased EEG microstate duration and anteriorisation of the brain electrical fields in mild and moderate dementia of the Alzheimer type psychiatry research. *Neuroimaging* 75, 183–191. doi: 10.1016/s0925-4927(97)00054-1
- Světlák, M., Bob, P., Erník, M., Chládek, J., and Kukleta, M. (2010). *Electrodermal dimensional complexity and smoking*. *Scripta Medica. Brno: Masarykova univerzita, roč.* 83, 63–69, 6.
- Thakur, G., Brevdo, E., Fučkar, N. S., and Wu, H. T. (2013). The Synchrosqueezing algorithm for time-varying spectral analysis: robustness properties and new paleoclimate applications. *Signal Process.* 93, 1079–1094. doi: 10.1016/j.sigpro.2012.11.029
- Thomas, K., Leslie, P., Lehmann, D., Pedro, V. S., Elisabeth, B., Horst, K., et al. (2002). Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. *Neuroimage* 16, 41–48. doi: 10.1006/nimg.2002.1070
- Totev, T., Taskov, T., and Dushanova, J. (2023). A wireless EEG system for Neurofeedback training. *Appl. Sci.* 13:96. doi: 10.3390/app13010096
- Uyulan, C., Ergüzel, T. T., Unubol, H., Cebi, M., Sayar, G. H., Nezhad Asad, M., et al. (2020). Major depressive disorder classification based on different convolutional neural network models: deep learning approach. *Clin. EEG Neurosci.* 52, 38–51. doi: 10.1177/1550059420916634
- Wacker, M., and Witte, A. (2011). The matched Gabor transform—a tool for adaptive phase extraction. *Front. Comput. Neurosci.* 5:00223. doi: 10.3389/conf.fncom.2011.53.00223
- Walther, T. G. (2005). Cluster validation by prediction strength. *J. Comput. Graph. Stat.* 3, 511–528. doi: 10.1198/106186005X59243
- Wei, T., Taolin, C., and Xiaoqi, H. (2017). Background versus event craving: differentiating the different pathways of psychological craving for nicotine addiction. *Adv. Psychol. Sci.* 11, 1932–1941. doi: 10.3724/SPJ.1042.2017.01932
- Weifeng, L., Xiaoming, L., Ruomeng, D., and Xiaoying, T. (2017). Exploring differences between left and right hand motor imagery via spatio-temporal EEG microstate. *Computer Assist Surg* 22, 258–266. doi: 10.1080/24699322.2017.1389404
- World Health Organization (2009). International statistical classification of diseases and related health problems (the) ICD-10. *Acta Chir. Iugosl.* 56, 65–69. doi: 10.2298/ACI0903065V
- Yan, Y., Dahmani, L., Ren, J., Shen, L., Peng, X., Wang, R., et al. (2020). Reconstructing lost BOLD signal in individual participants using deep machine learning. *Nat. Commun.* 11:5046. doi: 10.1038/s41467-020-18823-9
- Yan Xue, X., Jia Hui, D., YaYun, C., LiBo, Z., and Ping, W. (2017). Effect of selective inhibition of reactivated nicotine-associated memories with propranolol on nicotine craving. *JAMA Psychiat.* 74, 224–232. doi: 10.1001/jamapsychiatry.2016.3907
- Yang, W. C., Zhang, P. L., Wu, D. H., and Xin, Z. (2013). A method of false component discriminant of EMD based on Kolmogorov-Smirnov test. *Appl. Mech. Mater.* 427–429, 2005–2008. doi: 10.4028/www.scientific.net/AMM.427-429.2005
- You, S., Lei, B., Wang, S., Chui, C. K., Cheung, A. C., Liu, Y., et al. (2022). Fine perceptive GANs for brain MR image super-resolution in wavelet domain. *IEEE Trans Neural Netw Learn Syst.* 4, 1–13. doi: 10.1109/TNNLS.2022.3153088
- Zernig, G., Ahmed, S. H., Cardinal, R. N., Morgan, D., Acquas, E., Foltin, R. W., et al. (2007). Explaining the escalation of drug use in substance dependence: models and appropriate animal laboratory tests. *Pharmacology* 80, 65–119. doi: 10.1159/000103923



OPEN ACCESS

EDITED BY

Shuqiang Wang,
Chinese Academy of Sciences (CAS), China

REVIEWED BY

Chao Wang,
Shenzhen University, China
Naifu Jiang,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE

Richu Jin
✉ jinrc@sustech.edu.cn
Jiang Liu
✉ liuj@sustech.edu.cn

RECEIVED 22 March 2023

ACCEPTED 09 May 2023

PUBLISHED 26 May 2023

CITATION

Jin R, Cai Y, Zhang S, Yang T, Feng H, Jiang H, Zhang X, Hu Y and Liu J (2023) Computational approaches for the reconstruction of optic nerve fibers along the visual pathway from medical images: a comprehensive review. *Front. Neurosci.* 17:1191999. doi: 10.3389/fnins.2023.1191999

COPYRIGHT

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

Computational approaches for the reconstruction of optic nerve fibers along the visual pathway from medical images: a comprehensive review

Richu Jin^{1,2*}, Yongning Cai¹, Shiyang Zhang², Ting Yang², Haibo Feng², Hongyang Jiang², Xiaoqing Zhang², Yan Hu² and Jiang Liu^{1,2,3*}

¹Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen, China, ²Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China, ³Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China

Optic nerve fibers in the visual pathway play significant roles in vision formation. Damages of optic nerve fibers are biomarkers for the diagnosis of various ophthalmological and neurological diseases; also, there is a need to prevent the optic nerve fibers from getting damaged in neurosurgery and radiation therapy. Reconstruction of optic nerve fibers from medical images can facilitate all these clinical applications. Although many computational methods are developed for the reconstruction of optic nerve fibers, a comprehensive review of these methods is still lacking. This paper described both the two strategies for optic nerve fiber reconstruction applied in existing studies, i.e., image segmentation and fiber tracking. In comparison to image segmentation, fiber tracking can delineate more detailed structures of optic nerve fibers. For each strategy, both conventional and AI-based approaches were introduced, and the latter usually demonstrates better performance than the former. From the review, we concluded that AI-based methods are the trend for optic nerve fiber reconstruction and some new techniques like generative AI can help address the current challenges in optic nerve fiber reconstruction.

KEYWORDS

optic nerve fiber, visual pathway, image segmentation, fiber tracking, artificial intelligence, medical image analysis

1. Introduction

The visual pathway is a general name for a series of brain tissues including the optic nerve (ON), optic chiasm (OC), optic tract (OT), lateral geniculate nucleus (LGN), optic radiation (OR), and visual cortex (VC) (Smith and Strottmann, 2001; Jäger, 2005). In the visual pathway, ON, OC and OT are formed by the axons of the retinal ganglion cells (Becker et al., 2010), while OR is formed by another type of optic nerve fibers. These two types of optic nerve fibers are connected at the LGN, which is a relay station of optic signals (Fujita et al., 2001). The optic nerve fibers along the visual pathway are responsible for the conduction of optic signals from the retina to the visual cortex and play significant roles in vision formation.

Optic nerve fibers can be affected by various ophthalmological diseases, e.g., glaucoma (Hernowo et al., 2011; Tellouck et al., 2016; Haykal et al., 2022), age-related macular degeneration (Prins et al., 2016; Yoshimine et al., 2018) and optic neuritis (Yamamoto et al., 2005; Spierer et al., 2010; Zhao et al., 2018), and neurological diseases, e.g., multiple sclerosis (MS) and Alzheimer's disease (AD) (Reich et al., 2010; Klistorner et al., 2015; Mutlu et al., 2018; Wang et al., 2021). For different types of diseases, optic nerve fibers would represent varied symptoms like edema, demyelination, atrophy and degeneration at different locations along the visual pathway, which would change the original morphological and even structural characteristics of the optic nerve fibers. Also, the severity of the symptoms is highly relevant to disease progression. Dysfunction of optic nerve fibers would cause serious vision problems; describing the status of optic nerve fibers in morphology and structure can help determine a patient's condition and choose the appropriate treatment strategy.

In addition, it is not uncommon that optic nerve fibers get compressed or damaged due to tumors and traumas (Romano et al., 2007, 2009; Chamberland et al., 2018). It requires a clear delineation of the morphological and structural status of the optic nerve fibers to evaluate the damage. Meanwhile, the accurate locations of the optic nerve fibers in the brain play significant roles in neurosurgery for compression release and damage repair. The location information of the optic nerve fibers is also crucial for radiation therapy to protect the optic nerve fibers from radiation (Isambert et al., 2008; Dai et al., 2021).

Currently, there are several imaging techniques that can provide an in-vivo delineation of the optic nerve fiber in the visual pathway. Particularly, computed tomography (CT) and magnetic resonance imaging (MRI) are used to reveal the optic nerve fibers at the anterior visual pathway, i.e., from the optic disc to the LGN (Tamraz et al., 1999; Wichmann and Müller-Forell, 2004), while diffusion tensor imaging (DTI) is usually applied to delineate the optic nerve fibers at the OR (Dayan et al., 2015b). These imaging techniques make it possible to evaluate the morphological and structural status of the optic nerve fibers and target their locations in the brain via in-vivo approaches, and reconstructing the optic nerve fibers from medical images can further facilitate these approaches.

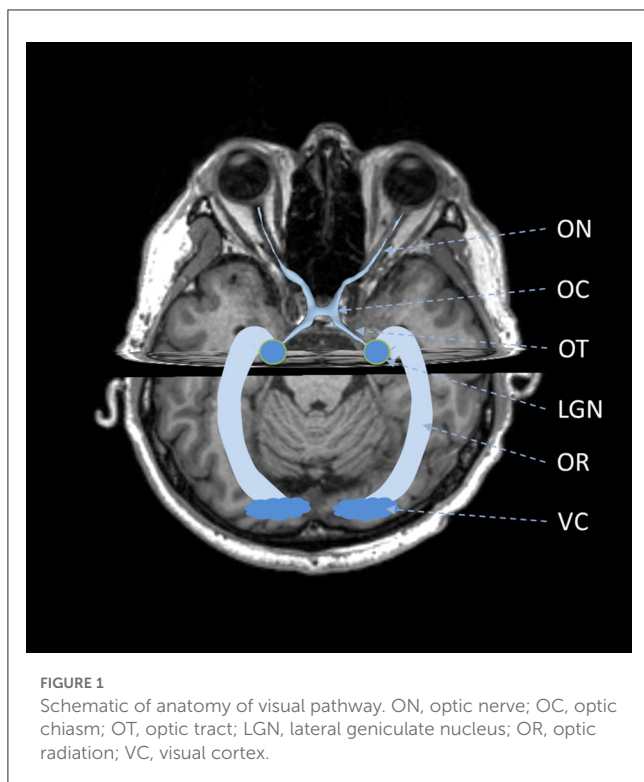
Manual optic nerve fiber reconstruction is difficult and time-consuming, thus computational approaches for automated optic nerve fiber reconstruction are developed. These computational approaches can be divided into two categories, i.e., image segmentation and fiber tracking. The former is used for CT/MRI images, while the latter is performed for DTI data. Despite the difference in implementation, these approaches face the same challenge, i.e., the thin-long structure of the optic nerve fibers. The thin-long structure makes the optic nerve fibers easily affected by the partial volume effect (PVE) (Mansoor et al., 2016). PVE can decrease the image contrast to neighboring tissues, increasing the difficulty of image segmentation (Cabezas et al., 2011); also, it enables multi-orientations in each voxel, raising the complexity of orientation estimation for fiber tracking (Alexander et al., 2001; Jeurissen et al., 2019). Though various computational approaches are proposed for this challenge in optic nerve fiber reconstruction, it has not been well addressed yet.

In recent years, some advanced techniques such as generative artificial intelligence (AI) have been developed and these techniques exhibit their potential in handling this challenge. Generative AI has demonstrated its power in image super-resolution and multi-modal image synthesis (Hu et al., 2020a,b, 2021; You et al., 2022). The major cause of PVE is the low image resolution, thus higher image resolution can help get it alleviated. Multi-modal image fusion is another way to resist PVE. Multi-modal images can provide consistent and complementary information to release the confusion caused by PVE. However, it is not common to see multi-modal approaches for optic nerve fiber reconstruction as the acquisition of multi-modal data would be expensive and time-consuming in clinical practice. Multi-modal image synthesis provides a cheap and efficient way to acquire multi-modal images (Hu et al., 2021), removing the biggest barrier that hinders multi-modal research on optic nerve fiber reconstruction.

To apply generative AI and other new techniques in optic nerve fiber reconstruction, it is better to gain a comprehensive understanding of the task and the existing methods. However, to the best of our knowledge, a comprehensive review of the computational approaches for the reconstruction of optic nerve fibers from medical images is still lacking. Therefore, we performed such a review in this paper. We started with the anatomy of the visual pathway and imaging techniques of the optic nerve fibers. Then, we described both the two strategies, i.e., image segmentation and fiber tracking, for optic nerve fiber reconstruction. For each strategy, both conventional and AI-based methods were introduced. Finally, we discussed the selection rules and future challenges to performing optic nerve fiber reconstruction, providing guidance for clinical application and future studies. More details can be viewed in the following sections.

2. Anatomy of visual pathway

The visual pathway consists of the ON, OC, OT, LGN, OR, and VC (Tamraz et al., 1999; Smith and Strottmann, 2001; Wichmann and Müller-Forell, 2004; Jäger, 2005), as shown in Figure 1. The ON is the first part of the visual pathway. It is a thin-long myelinated fiber bundle formed by the axons of the retinal ganglion cells. There is a pair of ONs, which start from the optic disks of each eye, pass through the orbit and optic canals, and finally get crossed at the OC. Based on the locations, the ON can be further divided into four segments, i.e., the intraocular, intraorbital, intracanalicular, and intracranial segments. The lengths for the four segments are about 1 mm, 30 mm, 6 mm, and 10 mm, respectively. The OC is a flat x-shape structure located at the junction of the floor and the anterior wall of the third ventricle. It is just situated anteriorly to the pituitary stalk. In OC, only the optic nerve fibers from the medial retina (nasal side) would get crossed, while those from the lateral retina (temporal side) remain uncrossed. Then, the optic nerve fibers at each side keep going from the posterolateral angle of the OC and form the left and right OTs. The optic nerve fibers in OTs run backward and lateralward of the OC and wind around the midbrain. Most of these optic nerve fibers get terminated at the LGN, while there are also some passing over the LGN and reaching the superior colliculus to coordinate eye movements. The LGN is located in the lateral geniculate body, which is the posterior-inferior



aspect of the thalamus. The LGN consists of alternating gray and white matter layers and serves as a relay station in the visual pathway (Fujita et al., 2001). The LGN projects the visual signal from the retina to the VC, and the optic nerve fiber connecting the LGN and VC form the optic radiation. The OR can be divided into three major fiber bundles, i.e., the dorsal, lateral, and ventral bundles. The dorsal and lateral bundles pass through the temporal and parietal lobes posteriorly and terminate at the occipital lobe; the ventral bundle runs anteriorly and laterally into the temporal lobe and bypasses the temporal horn of the lateral ventricle, generating the Meyer's loop (Tamraz et al., 1999; Dayan et al., 2015b). The VC is also called the striate cortex. It is located at the superior and inferior lips of the calcarine fissure.

It is seen that the major components of the visual pathway are optic nerve fibers. Separated by the LGN, the two types of optic nerve fibers share similar symmetric curved thin-long structures but vary from each other in length, diameter, and curvature. Also, they are located with different neighboring tissues. The optic nerve fibers in the anterior visual pathway, i.e., from the optic disc to the LGN, are mostly located around muscle, fat, cranium, and blood vessels, while those in the OR are next to the brain's gray and white matters. The differences in these anatomical characteristics lead to quite different representations in medical images. More details on imaging would be introduced in the next section.

3. Imaging of optic nerve fibers in visual pathway

Currently, CT, MRI and DTI are the common imaging techniques for in-vivo delineation of the optic nerve fibers in the

visual pathway. CT and MRI images are usually used to assess the optic nerve fibers in the anterior visual pathway (Smith and Strottmann, 2001; Becker et al., 2010). In general, MRI is superior in imaging the optic nerve fibers than CT as it can achieve high contrast among soft tissues. In comparison with CT, it can more easily differentiate the optic nerve fibers from the complex adjacent tissues in the orbit and sellar regions. Also, MRI is free from radiation and is safer than CT. Nevertheless, CT has its advantage in revealing bony tissues and foreign bodies. It can reveal the damages to the visual pathway caused by orbital or optic canal trauma as well as the calcification of the optic nerve fibers due to the tumor compression. In addition, CT is less affected by motion artifacts and can be applied to people with metal implants, making it a better choice than MRI in some special clinical scenarios. Besides CT and MRI, DTI can also be applied to reconstruct the optic nerve fibers in the anterior visual pathway; and, it can provide more details such as the fiber crossing at the OC (He et al., 2021). But, DTI takes much longer scanning time than MRI and CT, making it less practical in clinical scenarios. Instead, DTI is more frequently applied to reconstruct the optic nerve fibers in the OR (Dayan et al., 2015a; Schurr et al., 2018). The three fiber bundles in OR are located very close to other white matter tracts; the subtle variations in white matter signal make it difficult to reveal the anatomical heterogeneity in OR from CT and conventional MR images (Yogarajah et al., 2009; Winston et al., 2012). DTI is a technique to monitor the motion of water molecules in the human body by collecting multi-gradient MRI images. As the motion of water molecules is bounded by the nerve fibers, its speed and direction can be used to describe the structure and orientation of nerve fibers. Unlike MRI and CT images, the structure of optic radiation cannot be directly viewed in raw DTI images. There is a need to calculate the DTI metrics or perform fiber tracking to reveal the structure of the OR. It is seen that the three imaging techniques have their unique advantages and their own application scenarios. Also, the different representations of the optic nerve fibers in images of different modalities require different reconstruction methods. Usually, image segmentation is applied to CT and MRI images where optic nerve fibers exhibit a certain image contrast to neighboring tissues, while fiber tracking is performed to DTI data to exploit the structural and orientational information for more precise delineation of the optic nerve fibers. Both image segmentation and fiber tracking approaches can be further classified as conventional and AI-based methods. In the following two sections, we would describe more details of both the conventional and AI-based methods using the two reconstruction strategies.

4. Fiber reconstruction by image segmentation

4.1. Conventional methods

Image segmentation is usually used to reconstruct optic nerve fibers from CT and MRI images. Conventionally, there are various methods to perform image segmentation, such as thresholding, boundary-based, region-based, model-based, atlas-based, etc (Despotović et al., 2015; Wang et al., 2016). The

thresholding methods are not suitable for the segmentation of the optic nerve fibers given their poor image contrast with neighboring tissues at some segments of the visual pathway. Also, their thin-long structures make it difficult to perform boundary-based and region-based methods. It is found that most methods for optic nerve fiber reconstruction are model-based or atlas-based (Table 1).

Model-based methods would first define a model based on the prior information on the shape and appearance of the tissue to be segmented and then fit the model to the new images (Heimann and Meinzer, 2009). The models can be either fixed geometry models or deformable models. For the fixed geometry models, they can be easily fitted via an explicit parameter estimation based on selected landmarks. The deformable models such as active shape models, active appearance models and level-set are usually fitted with searching or optimization procedures. Particularly, Bekes et al. (2008) approximate the ON in a CT image as a cone and cylinder and fit the cone and cylinder using a semi-automatic way. This fixed-model-based approach is simple but its reproducibility is doubted. Noble and Dawant (2011) applied an atlas-navigated optimal medial axis and deformable model (NOMAD) to segment the ON and OC based on paired CT and T1-weighted MRI images. The exploitation of multi-modal images and hybrid methods (model- and atlas-based) enhances the segmentation results, but the paired CT and MRI images are not always available in clinical practice. Yang et al. (2014) proposed a weighted partitioned active shape model to segment the anterior visual pathway from T1-weighted MRI images. This method is also capable to segment the OT, which is believed as a more challenging task than ON and OC segmentation before this study. Mansoor et al. (2015) developed a method entitled PARTitioned Shape and Appearance Learning (PASCAL) to segment the anterior visual pathway from MRI images. This method can also be applied to the pathological anterior visual pathway.

Atlas-based methods treat the segmentation problem as a registration problem, i.e., aligning the new image and the atlas (Cabezas et al., 2011). Usually, an atlas contains two image volumes, one intensity image (template) and one segmented image (label). Image registration is used to build the geometrical connection between the new image and the template; then, the segmentation can be achieved by propagating the label to the image space via the geometrical connection. D'haese et al. (2003) manually drew an atlas that includes the ON based on visually selected MRI images and segmented the ON with the atlas. Gensheimer et al. (2007) extended single-atlas segmentation to multi-atlas segmentation and performed additional post-processing procedures including a ray casting algorithm, reshaping of unreasonable cross sections, and surface fitting to further modify the inaccurate contours. Isambert et al. (2008) applied a multi-atlas segmentation method to segment ON and OC from MRI images under clinical conditions. Asman et al. (2013) developed a non-local model to perform multi-atlas segmentation for the ON based on CT images. Harrigan et al. (2014) and Panda et al. (2014) paid attention to the robustness of the atlas-based segmentation for the ON and proposed an improved registration procedure.

4.2. AI-based methods

AI-based methods are data-driven approaches, which learn the rules from the data. Such approaches reduce manual operations like predefining models or atlases and are more easily implemented in practice. AI-based methods usually treat the segmentation procedure as a pixel/voxel-wise classification or clustering task. In the beginning, the classification/clustering is performed using conventional machine learning algorithms based on hand-crafted features. For example, Dolz et al. (2015) extracted features from neighborhood information and applied the support vector machine (SVM) to conduct the classification. With the occurrence and development of deep learning techniques, it becomes possible to integrate the feature extraction procedure into the learning process, further simplifying the procedure to segment the optic nerve fibers.

The studies on deep learning methods for optic nerve fiber segmentation from CT/MRI images have passed through three periods (Table 2). In the early period, deep learning methods are only used for feature extraction and segmentation is still implemented by conventional methods. For example, Mansoor et al. (2016) used a stacked auto-encoder to learn new feature representations for a model-based segmentation procedure. After this early period, deep learning is also used for pixel/voxel classification. At this stage, the network is usually formed by two network modules, e.g., a convolutional neural network (CNN) and a fully connected network, responsible for feature extraction and pixel/voxel classification, respectively. Based on this basic network structure, Ren et al. (2018) extended the original CNN to an interleaved structure for joint segmentation of optic nerve and chiasm; Dolz et al. (2017) replaced the CNN with a stacked denoised auto-encoders to learn a compact representation of the hand-crafted features; Duanmu et al. (2020) modified the CNN using a multi-resolution path approach to combine multi-scale features. Recently, a more powerful network, i.e., the U-Net, has been developed (Ronneberger et al., 2015). U-Net is composed of a down-sampling branch and an up-sampling branch. The down-sampling and up-sampling branches are made up of paired encoders and decoders, respectively. The down-sampling procedure can help extract the context information and the up-sampling procedure is used for fine localization. Also, there are skip connections between the encoders and decoders. As there might be information loss during the up-sampling procedure, the skip connections make it possible to combine the up-sampling results with the original information before the down-sampling procedure. With the skip connections, the localization can be more accurate. Compared with the two-module network, U-Net further integrates the feature extraction and pixel/voxel classification procedures.

The state-of-the-art (SOTA) methods for optic nerve fiber reconstruction from CT/MRI images are mostly based on the U-Net. Particularly, some researchers tried to modify the encoders and decoders as well as their connections to enhance context information exploitation. For example, Zhu et al. (2019) added squeeze-excitation blocks into the down-sampling and up-sampling approaches; Tong et al. (2019) and Zhu et al. (2021) tried DenseNet and V-Net, which enhance the connections among encoders and decoders, to segment the ON and OC from CT and MRI images. Also, some researchers tried to add

TABLE 1 Conventional image segmentation methods for optic nerve fiber reconstruction.

Method type	Method description	Anatomical region	Imaging modality	Research
Model-based	Geometry model	ON	CT	Bekes et al., 2008
	Atlas-navigated optimal medial axis and deformable model (NOMAD)	ON, OC	CT, MRI	Noble and Dawant, 2011
	Weighted partitioned active shape model	ON, OC, OT	MRI	Yang et al., 2014
	Partitioned Shape and Appearance Learning (PASCAL)	ON, OC, OT	MRI	Mansoor et al., 2015
Atlas-based	Single-atlas	ON	MRI	D'haese et al., 2003
	Multi-atlas, Post-processing	ON, OC	CT	Gensheimer et al., 2007
	Multi-atlas	ON, OC	MRI	Isambert et al., 2008
	Multi-atlas, non-local model	ON	CT	Asman et al., 2013
	Multi-atlas, variable voxel resolution and field of view	ON	CT	Harrigan et al., 2014; Panda et al., 2014

ON, optic nerve; OC, optic chiasm; OT, optic tract.

TABLE 2 AI-based image segmentation methods for optic nerve fiber reconstruction.

Method type	Method description	Anatomical region	Imaging modality	Dataset	Research
Machine Learning	SVM	ON	MRI	Private	Dolz et al., 2015
CNN only	Stacked auto-encoder	ON, OC, OT	MRI	Private	Mansoor et al., 2016
CNN+FCN	Stacked denoised auto-encoders+FCN	ON, OC	MRI	Private	Dolz et al., 2017
	Interleaved CNN+FCN	ON, OC	CT	PDCCA	Ren et al., 2018
	Multi-resolution multi-scale CNN+FCN	ON, OC	CT	Private	Duanmu et al., 2020
U-Net-Like	Squeeze-excitation Block	ON, OC	CT	PDDCA, TCIA	Zhu et al., 2019
	Connection Enhancement, Global restriction	ON, OC	CT	PDCCA	Tong et al., 2018
	Connection Enhancement, Global restriction	ON, OC	CT	PDCCA	Tong et al., 2019
	Connection Enhancement	ON, OC	CT	Private	Zhu et al., 2021
	Recursive ensemble segmentation	ON, OC	MRI	Private	Chen et al., 2019
	Localization+Segmentation	ON, OC	CT	PDCCA	Wang et al., 2019
	Localization+Segmentation, Atlas information	ON, OC, OT	MRI	Private	Zhao et al., 2019
	Localization+Segmentation, Atlas information	ON, OC, OT	CT, MRI	Private, PDCCA	Ai et al., 2020
	Localization+Segmentation	ON, OC	MRI	Private	Liu and Gu, 2020
	Localization+Segmentation	ON, OC	CT	PDCCA	Amjad et al., 2022
	Pre-processing	OC	MRI	CHIAS M	Puzniak et al., 2021b
	Pre-processing	ON	CT	TCIA	Ranjbarzadeh et al., 2022
	Post-processing	ON, OC	CT	Private	Ibragimov and Xing, 2017
	Post-processing	ON, OC	MRI	Private	Mlynarski et al., 2020

ON, optic nerve; OC, optic chiasm; OT, optic tract; SVM, support vector machine; CNN, convolutional neural networks.

global loss restrictions to avoid irregular segmentation results due to the pixel/voxel-wise segmentation strategy. Specifically, [Tong et al. \(2018, 2019\)](#) added a latent shape restriction as well as an adversarial restriction to guarantee the global shape of the segmented ON and OC. Besides the modification of network

blocks and losses, some researchers paid attention to the training strategies. [Chen et al.](#) proposed a recursive ensemble organ segmentation framework. In this framework, the organs that are easily segmented, e.g., the eyeballs, would be first segmented; and then, the segmentation results are fed to the network together

with the original inputs for the segmentation of more complicated organs like ON and OC (Chen et al., 2019). Wang et al. proposed a hybrid network containing two U-Nets for localization and segmentation, respectively. The U-Net for localization was named “LocNet” and used to localize the region of the ON, while the one for segmentation was named “SegNet” and applied only in the extracted region to exclude other interference (Wang et al., 2019). Zhao et al. adopted a similar strategy but replaced the LocNet with an atlas-based approach, i.e., performing registration between the atlas and a new image to localize the ON. They also generated a spatial probabilistic distribution map using the atlas to assist the segmentation (Zhao et al., 2019; Ai et al., 2020). Differently, Liu and Gu (2020) and Amjad et al. (2022) replaced the SegNet with a two-module network, where the CNN adopted a multi-resolution structure.

In addition to the deep learning networks, researchers also tried to enhance the segmentation results using proper pre-processing and post-processing approaches. For the pre-processing, Puzniak et al. (2021b) applied a data-augmentation strategy to train a 3D U-Net. Ranjbarzadeh et al. (2022) pre-processed the input images by combining a fuzzy C-mean clustering algorithm, histogram equalization, and a texture descriptor based on the local directional number. For post-processing, Ibragimov and Xing (2017) proposed a post-processing procedure based on Markov random fields. Mlynarski et al. (2020) developed a graph-based post-processing approach to guarantee the connectivity between the eyes and OC.

5. Fiber reconstruction by fiber tracking

5.1. Conventional methods

Fiber tracking, also called fiber tractography, is a computational procedure to reconstruct nerve fibers from DTI images. Although there is a debate on the reliability of fiber tracking in delineating the true brain nerve fibers, it has been widely applied in both medical research and clinical practice. There are also plenty of studies focusing on the reconstruction of optic nerve fibers, especially for the OR, using fiber tracking.

Fiber tracking would estimate a series of streamlines to delineate the global fiber tractography using deterministic, probabilistic, or global algorithms (Jeurissen et al., 2019; Li et al., 2020). Deterministic algorithms are proposed based on the assumption that there is a predominant orientation in each voxel of DTI images. Common deterministic algorithms include streamlines tracking (STT) (Basser, 1998; Basser et al., 2000), fiber assignment by continuous tracking (FACT) (Mori et al., 1999; Chao et al., 2008), Tensor-lines (Weinstein et al., 1999), tensor deflection (TEND) (Lazar et al., 2003), and vector criterion tracking (VCT) (Kim et al., 2004). These algorithms usually select the diffusion tensor as the model to describe fibers' microstructures at each voxel. But, the diffusion signal would be inevitably distorted by noise and artifacts, affecting the certainty of voxel orientation inferred from the diffusion tensor (Jones, 2010). The assumption of one orientation per voxel is also doubted due to the existence of crossing fibers (Behrens et al., 2007). The existence of these

problems raises concerns about the deterministic algorithms; the probabilistic algorithms are then proposed. To cope with the uncertainty, the probabilistic algorithms use the probability density functions (PDF) (Behrens et al., 2003) and fiber orientation distribution (FOD) (Tournier et al., 2004) to represent fibers' microstructures at each voxel. Based on these probabilistic models, the algorithms like probabilistic index of connectivity (PICO) (Parker et al., 2003), unscented Kalman filter (UKF) (Malcolm et al., 2010), probabilistic tracking with crossing fibers (PROBTRACKX) (Behrens et al., 2007), ConTrack (Sherbondy et al., 2008), particle filtering tractography (PFT) (Zhang et al., 2009), and 2nd-order Integration over Fiber Orientation Distributions (iFOD2) (Smith et al., 2012) are proposed. Compared with deterministic algorithms, probabilistic algorithms can delineate more complicated nerve fiber distributions; but, they would also cause a large number of false positive streamlines and suffer from heavy computational costs. Both the deterministic and probabilistic algorithms are based on local information, while global algorithms treat fiber tracking as a global optimization problem. The existing global algorithms can be mostly divided into two categories, i.e., graph-based algorithms (Iturria-Medina et al., 2007) and Gibbs algorithms. Graph-based algorithms should set the seeding and targeting regions, which is not necessary for Gibbs algorithms (Kreher et al., 2008). Global algorithms can avoid the error accumulation problem in local algorithms and reduce the number of false positive streamlines; but, their computational costs are much greater than local algorithms and convergent solutions are not guaranteed.

Besides the algorithm, there are also some key operations and settings to ensure an accurate fiber tracking procedure (Jacquesson et al., 2019; Jeurissen et al., 2019). For local algorithms and graph-based global algorithms, there is a need to determine the seeding and target regions of interest (ROIs). The seeding and target ROIs mean the two ends of the generated fibers by the tracking algorithms. Except for the whole-brain tracking, these two ROIs can be drawn in a manual way (Rossi-Espagnet et al., 2020; Haykal et al., 2022) or by projecting the labels in a built brain atlas (Karahan et al., 2019; Papadopoulos et al., 2021). In addition, the ROIs can also be acquired by other fiber tracking procedures (Davion et al., 2020). Besides these two types of ROIs, there are also inclusive and exclusive ROIs for the filtering of valid fibers (Horbrugger et al., 2019). In addition, some thresholds to constrain the fibers' lengths, curvatures/angles, and fractional anisotropy (FA)/fiber orientation distribution function (fODF) values are also set for the filtering process.

The specific methods for optic nerve fiber reconstruction are shown in Table 3. The reconstruction of the optic nerve fibers from DTI images follows the above fiber tracking frameworks; but, the selection of tracking algorithms, ROI drawing, and thresholds setting would change with the location of optic nerve fibers. Particularly, deterministic algorithms can be applied to the optic nerve fibers in the anterior visual pathway (Dasenbrock et al., 2011; De Blank et al., 2013; Takemura et al., 2017; Hofstetter et al., 2019; Jin et al., 2019) but they are not suggested for OR reconstruction (Yogarajah et al., 2009). The OR region is close to the neighboring white matter tracts and image voxels in this region are more likely to contain multiple orientations. The probabilistic algorithms can

be applied for both the two types of optic nerve fibers (Dayan et al., 2015a; Kammen et al., 2016; Zolal et al., 2016; Backner et al., 2018; Yoshimine et al., 2018; Ather et al., 2019; Glick-Shames et al., 2019; Wu et al., 2019; Davion et al., 2020; Lacerda et al., 2020; Rossi-Espagnet et al., 2020; Reid et al., 2021; Haykal et al., 2022; Liu et al., 2022); but, there are still some differences. The probabilistic algorithms are proposed to handle the uncertainty and they can be classified into different categories based on the source of the uncertainty (Jeurissen et al., 2019). The reconstruction of optic nerve fibers at the anterior visual pathway and in the OR has different uncertainty sources. The former's uncertainty comes from the interference of the complicated skull base environment, which contains nerves, bone, air, soft tissue, and cerebrospinal fluid; the latter's uncertainty is mainly due to the multi-orientation problem. The difference in uncertainty sources would affect the selection of the probabilistic algorithms. In addition, the seeding ROIs for the ON and OT reconstruction are usually set as the end of the eyeballs and the OC, respectively, while those for the OR are set as the LGN. The target ROIs include the OC, LGN and VC for the reconstruction of the ON, OT and OR, respectively. The other settings like the inclusive and exclusive ROIs as well as the thresholds would be more task-specific.

5.2. AI-based methods

The conventional framework for fiber tracking is a complicated procedure containing the processes like pre-processing, seeding, tractography, and filtering of valid streamlines. Although several softwares integrate these processes (Table 3), the operations like ROI drawing, tracking algorithm selection, and threshold setting still require manual implementation. In recent years, AI technique has developed rapidly; researchers are trying to replace these manual operations with automated ways using AI technique (Table 4).

AI-based methods are applied first in the tractography process. Neher et al. (2017) tried to perform the tractography by machine learning. They applied a random forest classifier to learn multiple potential directions of a streamline from the raw diffusion signals and determined the streamline's progressing direction and termination using a neighborhood sampling strategy and a voting scheme, respectively. Poulin et al. treated the tractography as a regression problem and proposed the recurrent neural networks (RNN) to acquire the mapping between the diffusion signal and the streamlines' directions for both whole-brain and bundle-specific tractography (Poulin et al., 2017, 2018). The RNN can exploit both the new observations and the past seen information along the tracked streamlines. In addition to the diffusion signals, Jörgens et al. (2018) further pointed out the importance of the previous step directions for the tractography. They adopted an alternative way to predict the next step direction of a streamline via a multi-layer perceptron (MLP), whose input is a vector acquired by concatenating the diffusion signals and previous step directions. Wegmayr et al. (2018) also used an MLP to perform the tractography and further validated the significance of previous step directions; but, they changed the input of the

MLP as a vector formed by a flattened data block and several incoming vectors. The tractography can also be implemented via reinforcement learning and Théberge et al. (2021) proposed a general framework for this strategy. Apart from these local tractography methods, Wasserthal et al. (2018) developed a U-Net-like network to directly reconstruct the fiber tracts from the fields of fODF peaks.

Recently, several AI-based methods have been applied in processes other than tractography. For the pre-processing approach, AI-based methods focus on two aspects, i.e., generating high-fidelity diffusion signals from low-quality input and building the diffusion model from the raw diffusion signals. Acquiring high-fidelity diffusion signals usually requires a certain number of diffusion-encoding directions and multi-shell acquisitions, which takes a long scanning time. Tian et al. (2020) proposed a 10-layer CNN to reduce the requirement on the number of diffusion-encoding directions, particularly limiting the number to the minimum level for diffusion tensor calculation. Koppers et al. (2017) and Jha et al. (2022a) reconstructed the multi-shell diffusion signals from single-shell acquisitions using DNN and U-Net-like network, respectively. Zeng et al. (2022) proposed a super-resolution network to enhance the FOD model that was built based on the single-shell acquisition, and Jha et al. (2022b) developed a more complicated network containing multiple encoder-decoder structures and discriminators. Mapping raw diffusion signals to diffusion models is also very challenging. It is quite difficult for conventional methods to estimate the fibers' number and orientations per voxel from raw diffusion signals. Li et al. (2021) demonstrated the advantages of AI-based methods in this challenging task. They proposed a SuperDTI network for diffusion model generation and the test results suggest that their model is less sensitive to noise and more robust to misregistration than conventional tensor fitting methods. Karimi et al. (2021a,b,c) further verified the superiority of AI-based methods via a series of explorations on diffusion metric map generation, fODF generation and fibers' number and orientations estimation. In addition to the pre-processing process, AI-based methods are also used to achieve automatic seeding. Avital et al. (2019) and Wasserthal et al. (2019) tried automated seeding using U-Net and U-Net-like network, respectively. There are also studies focusing on AI-based automated filtering of valid streamlines. Particularly, AI can be used to draw inclusive or exclusive ROIs, such as He's work (He et al., 2023). Also, AI models can be used to directly classify or cluster the reconstructed streamlines (Xu et al., 2019; Zhang et al., 2020; Chen et al., 2021; Xue et al., 2022, 2023).

Although multiple AI-based methods are proposed for fiber tracking, the application of these methods in optic nerve fiber reconstruction is still rare. To the best of our knowledge, Reid et al. (2021) applied a U-Net-like network to automatically draw seeding ROI at the optic tract. He et al. (2023) proposed a unified global tractography framework for automatic visual pathway reconstruction. Li et al. (2022) used a modified SupWMA network to cluster the streamlines in the anterior visual pathways. These methods demonstrate the feasibility and effectiveness of AI-based methods in optic nerve fiber reconstruction, while there is still room for further improvement.

TABLE 3 Conventional fiber tracking approaches for optic nerve fiber reconstruction.

Method type	Model/algorithm	Anatomical region	Software	Research
Deterministic	Tensor/FACT	OT	DTI Studio	Dasenbrock et al., 2011
	Tensor/FACT	OT	TrackVis	Jin et al., 2019
	Tensor/FACT	OT	dTV II FZR _x	Takemura et al., 2017
	Tensor/FACT	ON, OT, OR	DTI Studio	De Blank et al., 2013
	FOD/STT	OT, OR	ExploreDTI	Hofstetter et al., 2019
Probabilistic	Tensor/PICo	ON	FSL	Zolal et al., 2016
	Tensor/PICo	OR	Camino	Dayan et al., 2015a
	Tensor/ConTrack	OT, OR	VISTA	Backner et al., 2018
	Tensor/ConTrack	OT, OR	VISTA	Glick-Shames et al., 2019
	Tensor/ConTrack	OT, OR	VISTA	Yoshimine et al., 2018
	Tensor/PROBTRACKX	ON, OT	FSL	Wu et al., 2019
	Tensor/PROBTRACKX2	ON, OC, OT	FSL	Ather et al., 2019
	FOD/iFOD2	OR	MRtrix3	Davion et al., 2020
	FOD/iFOD2	OR	MRtrix3	Lacerda et al., 2020
	FOD/iFOD2	OT, OR	MRtrix3	Rossi-Espagnet et al., 2020
	FOD/iFOD2	OR	MRtrix3	Reid et al., 2021
	FOD/iFOD2	OT, OR	MRtrix3	Haykal et al., 2022
	FOD/iFOD2	OR	MRtrix3	Liu et al., 2022

ON, optic nerve; OC, optic chiasm; OT, optic tract; OR, optic radiation; FOD, fiber orientation distribution.

6. Discussion

Optic nerve fiber reconstruction is a common step to evaluate or project optic nerve fibers in clinical diagnosis and treatment. As shown in section 2, optic nerve fibers have thin-long structures and varying curvatures at different segments of the visual pathway, making them difficult to evaluate in either qualitative or quantitative ways without the reconstruction from the medical images. Also, manual delineation of the optic nerve fibers would be a tough task and costs a lot of time. As a result, computational methods are highly needed for clinical applications on optic nerve fibers. It is found that optic nerve fibers can be revealed in images of multiple modalities and there are different reconstruction strategies for each imaging modality. Also, each reconstruction strategy has both conventional and AI-based implementations. This paper reviews the existing computational methods to guide optic nerve fiber reconstruction in medical research and clinical practice and demonstrates the trend for future studies.

CT and MRI images are widely used for the visualization of the optic nerve fibers at the anterior visual pathway, i.e., from the end of the eyeballs to the LGN, while DTI can be used to visualize the optic nerve fibers along the entire visual pathway. Even though, DTI would not replace CT and MRI for optic nerve fiber reconstruction in clinical practice at the current stage. On one hand, DTI is with longer scanning time and lower image resolution than CT and MRI, making it less applicable in clinical practice. On the other hand, there are still debates on the consistency between the reconstructed fibers from

DTI data and the real fibers in anatomy ([Jeurissen et al., 2019](#)), which limits its application scenarios such as the OAR drawing in radiation therapy.

Image segmentation and fiber tracking are two different reconstruction approaches for CT/MRI and DTI, respectively. Besides that, there are some other differences between these two approaches. Fiber tracking can achieve a more precise delineation of the optic nerve fibers than image segmentation, allowing the extraction of more accurate features to describe the morphological and structural changes of optic nerve fibers. For example, optic nerve fiber degeneration can be described by the volume change based on image segmentation results while it can be more precisely evaluated by the reduction in optic nerve fiber number based on fiber tracking results. Nevertheless, fiber tracking is time-consuming and its computational process is complicated and easily affected by noises and artifacts ([Tournier et al., 2002](#); [Lazar and Alexander, 2003](#)). Also, it is not uncommon that there are false positive results and it requires abundant experience and enough knowledge of brain anatomy to ensure an accurate result ([Jeurissen et al., 2019](#)). These drawbacks restrict the scenarios where it can apply in clinical practice. In comparison to fiber tracking, image segmentation would be more efficient and robust; also, its results can be easily evaluated.

In comparison with conventional methods, AI-based methods are believed to be the trend for both image segmentation and fiber tracking. For image segmentation, AI-based methods are preferred to conventional model-based and atlas-based methods. Model-based methods require the design of complicated

TABLE 4 AI-based fiber tracking approaches.

Process	AI model	Research
Pre-processing	CNN	Tian et al., 2020
	DNN	Koppers et al., 2017
	U-Net-like	Jha et al., 2022a
	SRN	Zeng et al., 2022
	Encoder-decoder, Discriminator	Jha et al., 2022b
	SuperDTI network	Li et al., 2021
	MLP, U-Net	Karimi et al., 2021a,b,c
Seeding	U-Net	Avital et al., 2019
	U-Net-like	Wasserthal et al., 2019
Tractography	Random forest classifier	Neher et al., 2017
	RNN	Poulin et al., 2017, 2018
	MLP	Jörgens et al., 2018;
	MLP	Wegmayr et al., 2018
	Reinforcement learning	Théberge et al., 2021
	U-Net-like	Wasserthal et al., 2018
Filtering	U-Net-like	He et al., 2023
	CNN	Xu et al., 2019
	CNN	Zhang et al., 2020
	Siamese networks	Chen et al., 2021
	Contrast learning	Xue et al., 2022, 2023

CNN, convolutional neural networks; DNN, deep neural networks; RNN, recurrent neural networks; SRN, super-resolution networks.

models to fit the thin-long structure of the optic nerve fiber; such models are difficult to estimate based on the complex background along the visual pathway and their robustness is doubted. Atlas-based methods require the registration between the target and template, while it is not easy to get two images fully aligned given the individual differences and interference from noises and artifacts. AI-based methods are data-driven approaches, which can automatically learn rules from complicated data. AI-based methods are more easily performed than those conventional methods and demonstrate much better segmentation accuracy and robustness. The only disadvantage of AI-based methods now is their high demand for fine-annotated labels.

For fiber tracking, the superiority of AI-based methods over conventional methods is not as great as image segmentation at the current stage. On one hand, AI-based methods are mostly proposed for one certain step of the fiber tracking procedure and a proper end-to-end AI-based fiber tracking approach is still lacking. On the other hand, the conventional methods for each fiber tracking step have been well integrated into toolboxes and software, decreasing their difficulty in implementation. Even though, it is seen that more and more studies on fiber tracking are trying to replace the conventional methods with AI-based ones.

It is also noticed that there are still some challenges in the reconstruction of optic nerve fibers from medical images with AI-based methods. These challenges point out the direction of future studies. The first challenge is the thin-long structure of the optic nerve fibers. The long optic nerve fibers pass through various brain regions that are formed by different brain tissues, yielding complicated contextual information. Meanwhile, the thin structure makes the signal intensities of the optic nerve fibers easily affected by their neighboring tissues due to the PVE, yielding varied signal intensities at different segments of the visual pathway. The existing image segmentation methods applied multi-scale, coarse-to-fine, or iterative strategies to handle the variations in signal intensity and contextual information; pre-processing and post-processing are also used to modify the false-positive and missing voxels. Even though measures are taken, it is seen that the improvement is far from satisfactory, suggesting that the current local voxel-based segmentation strategy would not be powerful enough to handle such a complicated problem. Also, the long optic nerve fibers have varied curvatures. In existing fiber tracking frameworks, the curvature is a significant sign for tracking termination and fiber selection. The varied curvatures increase the difficulty of setting these rules. Furthermore, it requires a large field of view to reveal the long optic nerve fibers in an image at the current stage. To achieve such a field of view, the image resolution has to be sacrificed to maintain an acceptable scanning time in clinical practice, increasing the PVE. Thus, more powerful segmentation and fiber tracking strategies are required to cope with the challenges brought by the thin-long structure.

The second challenge is the lacking of task-specific datasets. AI-based methods are data-driven methods and their performance highly depends on the quality of data. To the best of our knowledge, most existing studies on image segmentation and fiber tracking are based on private datasets, which are not available to the public. There are a minor number of publicly available datasets, such as PDDCA ([Raudaschl et al., 2017](#)), TCIA ([Clark et al., 2013](#); [Zhu et al., 2019](#)), and CHIASM ([Puzniak et al., 2021a,b](#)) for image segmentation and HCP for fiber tracking. However, these datasets are not initially collected for optic nerve fiber reconstruction. Most of these datasets require further cleaning and annotation operations. Meanwhile, the imaging protocols and pre-processing steps in these datasets may not be consistent with those used in clinical practice. Also, in some of the datasets, the images only cover part of the optic nerve fibers and cannot be used to reconstruct the entire visual pathway. In addition, the situations like multi-modal images and disease-specific deformation are not fully considered in these existing datasets. Therefore, building a dataset specifically for optic nerve fiber reconstruction is in great need.

The third challenge is the control of computational cost. For image segmentation, more powerful segmentation networks are usually with more complicated network structures at the current stage. Also, the inputs are 3D brain images for the reconstruction of optic nerve fibers. These together indicate a high computational cost. For fiber tracking, the tractography is usually an iterative process and time-consuming for both conventional and AI-based methods. The

high computational cost would reduce the value of clinical application. The way to balance the computational cost and reconstruction performance would be another challenge in future studies.

There are some new techniques such as generative AI that can help address these challenges. Generative AI has demonstrated its power in image super-resolution and image synthesis. Image super-resolution can be used to cope with the low-resolution problem caused by the large field of view. Also, image synthesis can be used to generate more data to get full exploitation of the existing datasets. In addition, it is realized that multi-modal fusion would be a possible way to enhance the performance of optic nerve fiber reconstruction. There are many other examples to support its effect on segmentation (Menze et al., 2014; Ibtehaz and Rahman, 2020; Wang et al., 2022). The combination of segmentation results and fiber tracking has once been explored (Reid et al., 2021; He et al., 2023). The segmentation results can be used as the seeds for tractography or the masks to filter valid streamlines. Therefore, developing new fusion and combination methods would be a feasible way to improve the reconstruction performance. Nevertheless, this kind of method would face the problem that multi-modal images are difficult to acquire in clinical practice. Generative AI provides a way for multi-modal image synthesis. Thus, in the future, we can try these new techniques in optic nerve fiber reconstruction.

7. Conclusion

In this paper, we provided a comprehensive review of the current SOTA computational methods for the reconstruction of optic nerve fibers. We described the difficulties to delineate or evaluate the optic nerve fibers directly from medical images, suggesting the necessity of optic nerve fiber reconstruction. We reviewed both the image segmentation and fiber tracking methods and the successful application of these methods in previous studies indicates the feasibility and effectiveness of computational methods in optic nerve fiber reconstruction. Also, we introduced both the conventional and AI-based implementations, and there is no doubt that AI-based methods are better choices for optic nerve fiber

reconstruction. Meanwhile, we also pointed out the challenges for the existing AI methods, and future studies are needed to address these challenges.

Author contributions

RJ and JL contributed to conception and design of the reviewing process. YC, SZ, TY, and HF searched and sorted the literatures. HJ, YH, and XZ provided key suggestions on the anatomy of visual pathway, imaging techniques, and image segmentation methods. RJ wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62101236 and 82102189), General Program of National Natural Science Foundation of China (Grant No. 82272086), Guangdong Provincial Department of Education (Grant No. 2020ZDZX3043), Guangdong Provincial Key Laboratory (Grant No. 2020B121201001), Shenzhen Natural Science Fund (JCYJ20200109140820699), and the Stable Support Plan Program (20200925174052004).

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

- Ai, D., Zhao, Z., Fan, J., Song, H., Qu, X., Xian, J., et al. (2020). Spatial probabilistic distribution map-based two-channel 3D U-net for visual pathway segmentation. *Pattern Recognit. Lett.* 138, 601–607. doi: 10.1016/j.patrec.2020.09.003
- Alexander, A. L., Hasan, K. M., Lazar, M., Tsuruda, J. S., and Parker, D. L. (2001). Analysis of partial volume effects in diffusion-tensor MRI. *Magn. Reson. Med.* 45, 770–780. doi: 10.1002/mrm.1105
- Amjad, A., Xu, J., Thill, D., Lawton, C., Hall, W., Awan, M. J., et al. (2022). General and custom deep learning autosegmentation models for organs in head and neck, abdomen, and male pelvis. *Med. Phys.* 49, 1686–1700. doi: 10.1002/mp.15507
- Asman, A. J., Delisi, M. P., Mawn, L. A., Galloway, R. L., and Landman, B. A. (2013). Robust Non-local Multi-atlas Segmentation of the Optic Nerve. *Medical Imaging 2013: Image Processing*. Bellingham, WA: SPIE. p. 434–441.
- Ather, S., Proudlock, F. A., Welton, T., Morgan, P. S., Sheth, V., Gottlob, I., et al. (2019). Aberrant visual pathway development in albinism: from retina to cortex. *Hum. Brain Mapp.* 40, 777–788. doi: 10.1002/hbm.24411
- Avital, I., Nelkenbaum, I., Tsarfaty, G., Konen, E., Kiryati, N., and Mayer, A. (2019). Neural segmentation of seeding ROIs (sROIs) for pre-surgical brain tractography. *IEEE Trans. Med. Imaging* 39, 1655–1667. doi: 10.1109/TMI.2019.2954477
- Backner, Y., Kuchling, J., Massarwa, S., Oberwahrenbrock, T., Finke, C., Bellmann-Strobl, J., et al. (2018). Anatomical wiring and functional networking changes in the visual system following optic neuritis. *JAMA Neurol.* 75, 287–295. doi: 10.1001/jamaneurol.2017.3880
- Basser, P. J. (1998). Fiber-tractography via diffusion tensor MRI (DT-MRI)“, in: *Proceedings of the 6th Annual Meeting ISMRM*. Sydney, Australia.
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* 44, 625–632. doi: 10.1002/1522-2594(200010)44:4%3C625::AID-MRM17%3E3.0.CO;2-O
- Becker, M., Masterson, K., Delavelle, J., Viallon, M., Vargas, M.-I., and Becker, C. D. (2010). Imaging of the optic nerve. *Eur. J. Radiol.* 74, 299–313. doi: 10.1016/j.ejrad.2009.09.029

- Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34, 144–155. doi: 10.1016/j.neuroimage.2006.09.018
- Behrens, T. E., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., et al. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088. doi: 10.1002/mrm.10609
- Bekes, G., Máté, E., Nyúl, L. G., Kuba, A., and Fidler, M. (2008). Geometrical model-based segmentation of the organs of sight on CT images. *Med. Phys.* 35, 735–743. doi: 10.1118/1.2826557
- Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., and Cuadra, M. B. (2011). A review of atlas-based segmentation for magnetic resonance brain images. *Comput. Methods Programs Biomed.* 104, e158–e177. doi: 10.1016/j.cmpb.2011.07.015
- Chamberland, M., Tax, C. M. W., and Jones, D. K. (2018). Meyer's loop tractography for image-guided surgery depends on imaging protocol and hardware. *Neuroimage-Clinical* 20, 458–465. doi: 10.1016/j.nicl.2018.08.021
- Chao, Y.-P., Chen, J.-H., Cho, K.-H., Yeh, C.-H., Chou, K.-H., and Lin, C.-P. (2008). A multiple streamline approach to high angular resolution diffusion tractography. *Med. Eng. Phys.* 30, 989–996. doi: 10.1016/j.medengphys.2008.01.010
- Chen, H., Lu, W., Chen, M., Zhou, L., Timmerman, R., Tu, D., et al. (2019). A recursive ensemble organ segmentation (REOS) framework: application in brain radiotherapy. *Phys. Med. Biol.* 64, 025015. doi: 10.1088/1361-6560/aaf83c
- Chen, Y., Zhang, C., Song, Y., Makris, N., Rathi, Y., Cai, W., et al. (2021). Deep fiber clustering: anatomically informed unsupervised deep learning for fast and effective white matter parcellation. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII*. Berlin: Springer. p. 497–507.
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* 26, 1045–1057. doi: 10.1007/s10278-013-9622-7
- Dai, X., Lei, Y., Wang, T., Zhou, J., Curran, W. J., Liu, T., et al. (2021). Deep attention mask regional convolutional neural network for head-and-neck MRI multi-organ auto-delineation. In: *Conference on Medical Imaging - Computer-Aided Diagnosis*, eds M. A. Mazurowski and K. Drukker (International Society for Optics and Photonics, SPIE), 115971B. doi: 10.1117/12.2581131
- Dasenbrock, H. H., Smith, S. A., Ozturk, A., Farrell, S. K., Calabresi, P. A., and Reich, D. S. (2011). Diffusion tensor imaging of the optic tracts in multiple sclerosis: association with retinal thinning and visual disability. *J. Neuroimage* 21, e41–e49. doi: 10.1111/j.1552-6569.2010.00468.x
- Davion, J.-B., Kuchcinski, G., Viard, R., Dumont, J., Pruvo, J.-P., Leclerc, X., et al. (2020). A fully automatic method for optic radiation tractography applicable to multiple sclerosis patients. *Brain Topogr.* 33, 533–544. doi: 10.1007/s10548-020-00771-8
- Dayan, M., Kreutzer, S., and Clark, C. A. (2015a). Tractography of the optic radiation: a repeatability and reproducibility study. *NMR Biomed.* 28, 423–431. doi: 10.1002/nbm.3266
- Dayan, M., Munoz, M., Jentschke, S., Chadwick, M. J., Cooper, J. M., Riney, K., et al. (2015b). Optic radiation structure and anatomy in the normally developing brain determined using diffusion MRI and tractography. *Brain Struct. Funct.* 220, 291–306. doi: 10.1007/s00429-013-0655-y
- De Blank, P. M. K., Berman, J. I., Liu, G. T., Roberts, T. P. L., and Fisher, M. J. (2013). Fractional anisotropy of the optic radiations is associated with visual acuity loss in optic pathway gliomas of neurofibromatosis type 1. *Neuro-oncology* 15, 1088–1095. doi: 10.1093/neuonc/not068
- Despotović, I., Goossens, B., and Philips, W. (2015). MRI segmentation of the human brain: challenges, methods, and applications. *Comput. Math. Methods Med.* 2015. doi: 10.1155/2015/450341
- D'haese, P.-F., Duay, V., Li, R., Du Bois D'aische, A., Merchant, T. E., et al. (2003). Automatic segmentation of brain structures for radiation therapy planning. In: *Medical Imaging 2003: Image Processing*. Bellingham, WA: SPIE. p. 517–526.
- Dolz, J., Leroy, H.-A., Reyns, N., Massotier, L., and Vermandel, M. (2015). A fast and fully automated approach to segment optic nerves on MRI and its application to radiosurgery. In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. Manhattan, NY: IEEE. p. 1102–1105.
- Dolz, J., Reyns, N., Betrouni, N., Kharroubi, D., Quidet, M., Massotier, L., et al. (2017). A deep learning classification scheme based on augmented-enhanced features to segment organs at risk on the optic region in brain cancer patients. *arXiv [Preprint]*. doi: 10.48550/arXiv.1703.10480
- Duanmu, H., Kim, J., Kanakaraj, P., Wang, A., Joshua, J., Kong, J., et al. (2020). Automatic brain organ segmentation with 3D fully convolutional neural network for radiation therapy treatment planning. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. Manhattan, NY: IEEE. p. 758–762.
- Fujita, N., Tanaka, H., Takanashi, M., Hirabuki, N., Abe, K., Yoshimura, H., et al. (2001). Lateral geniculate nucleus: anatomic and functional identification by use of MR imaging. *Am. J. Neuroradiol.* 22, 1719–1726. Available online at: <https://www.wofoscience.com/wos/allidb/full-record/WOS:000171752400019>
- Gensheimer, M., Cmelak, A., Niermann, K., and Dawant, B. M. (2007). Automatic delineation of the optic nerves and chiasm on CT images. In: *Medical Imaging 2007: Image Processing*. Bellingham, WA: SPIE. p. 406–415.
- Glick-Shames, H., Backner, Y., Bick, A., Raz, N., and Levin, N. (2019). The impact of localized grey matter damage on neighboring connectivity: Posterior cortical atrophy and the visual network. *Brain Imag. Behav.* 13, 1292–1301. doi: 10.1007/s11682-018-9952-7
- Harrigan, R. L., Panda, S., Asman, A. J., Nelson, K. M., Chaganti, S., Delisi, M. P., et al. (2014). Robust optic nerve segmentation on clinically acquired computed tomography. *J. Med. Imag.* 1, 034006–034006. doi: 10.1117/1.JMI.1.3.034006
- Haykal, S., Invernizzi, A., Carvalho, J., Jansonius, N., and Cornelissen, F. (2022). Microstructural visual pathway white matter alterations in primary open-angle glaucoma: a neurite orientation dispersion and density imaging study. *Am. J. Neuroradiol.* 43, 756–763. doi: 10.3174/ajnr.A7495
- He, J., Yao, S., Zeng, Q., Chen, J., Sang, T., Xie, L., et al. (2023). A unified global tractography framework for automatic visual pathway reconstruction. *NMR Biomed.* 34, e4904. doi: 10.1002/nbm.4904
- He, J., Zhang, F., Xie, G., Yao, S., Feng, Y., Bastos, D. C., et al. (2021). Comparison of multiple tractography methods for reconstruction of the retinogeniculate visual pathway using diffusion MRI. *Hum. Brain Mapp.* 42, 3887–3904. doi: 10.1002/hbm.25472
- Heimann, T., and Meinzer, H.-P. (2009). Statistical shape models for 3D medical image segmentation: a review. *Med. Image Anal.* 13, 543–563. doi: 10.1016/j.media.2009.05.004
- Hernowo, A. T., Boucard, C. C., Jansonius, N. M., Hooymans, J. M., and Cornelissen, F. W. (2011). Automated morphometry of the visual pathway in primary open-angle glaucoma. *Invest. Ophthalmol. Visual Sci.* 52, 2758–2766. doi: 10.1167/iovs.10-5682
- Hofstetter, S., Sabbah, N., Mohand-Said, S., Sahel, J.-A., Habas, C., Safran, A. B., et al. (2019). The development of white matter structural changes during the process of deterioration of the visual field. *Sci. Rep.* 9, 2085. doi: 10.1038/s41598-018-38430-5
- Horbrugger, M., Loewe, K., Kaufmann, J., Wagner, M., Schippling, S., Pawlitzki, M., et al. (2019). Anatomically constrained tractography facilitates biologically plausible fiber reconstruction of the optic radiation in multiple sclerosis. *NeuroImage Clin.* 22, 101740. doi: 10.1016/j.nicl.2019.101740
- Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., and Shen, Y. (2021). Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. *IEEE Trans. Med. Imaging* 41, 145–157. doi: 10.1109/TMI.2021.3107013
- Hu, S., Shen, Y., Wang, S., and Lei, B. (2020a). Brain MR to PET synthesis via bidirectional generative adversarial network. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020 Proceedings, Part II* 23. Berlin: Springer. p. 698–707.
- Hu, S., Yu, W., Chen, Z., and Wang, S. (2020b). Medical image reconstruction using generative adversarial network for Alzheimer disease assessment with class-imbalance problem. In: *2020 IEEE 6th International Conference on Computer and Communications*. Manhattan, NY: IEEE. p. 1323–1327.
- Ibragimov, B., and Xing, L. (2017). Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med. Phys.* 44, 547–557. doi: 10.1002/mp.12045
- Ibtehaz, N., and Rahman, M. S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural networks* 121, 74–87. doi: 10.1016/j.neunet.2019.08.025
- Isambert, A., Dhermain, F., Bidault, F., Commowick, O., Bondia, P.-Y., Malandain, G., et al. (2008). Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. *Radiother. Oncol.* 87, 93–99. doi: 10.1016/j.radonc.2007.11.030
- Iturria-Medina, Y., Canales-Rodríguez, E. J., Melie-García, L., Valdés-Hernández, P. A., Martínez-Montes, E., Alemán-Gómez, Y., et al. (2007). Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *Neuroimage* 36, 645–660. doi: 10.1016/j.neuroimage.2007.02.012
- Jacquesson, T., Frindel, C., Koccar, G., Berhouma, M., Jouanneau, E., Atty, A., et al. (2019). Overcoming challenges of cranial nerve tractography: a targeted review. *Neurosurgery* 84, 313–325. doi: 10.1093/neuros/nyy229
- Jäger, H. (2005). Loss of vision: imaging the visual pathways. *Eur. Radiol.* 15, 501–510. doi: 10.1007/s00330-004-2638-0
- Jeurissen, B., Descoteaux, M., Mori, S., and Leemans, A. (2019). Diffusion MRI fiber tractography of the brain. *NMR Biomed.* 32, e3785. doi: 10.1002/nbm.3785
- Jha, R. R., Jaswal, G., Bhavsar, A., and Nigam, A. (2022a). Single-shell to multi-shell dMRI transformation using spatial and volumetric multilevel hierarchical reconstruction framework. *Magn. Reson. Imaging* 87, 133–156. doi: 10.1016/j.mri.2021.12.011
- Jha, R. R., Pathak, S. K., Nath, V., Schneider, W., Kumar, B. R., Bhavsar, A., et al. (2022b). VRRNet: Volumetric ROI fODF reconstruction network for estimation of multi-tissue constrained spherical deconvolution with only single shell dMRI. *Magn. Reson. Imaging* 90, 1–16. doi: 10.1016/j.mri.2022.03.004

- Jin, Z., Bao, Y., Wang, Y., Li, Z., Zheng, X., Long, S., et al. (2019). Differences between generalized Q-sampling imaging and diffusion tensor imaging in visualization of crossing neural fibers in the brain. *Surg. Radiol. Anat.* 41, 1019–1028. doi: 10.1007/s00276-019-02264-1
- Jones, D. K. (2010). Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI. *Imaging Med.* 2, 341. doi: 10.2217/iim.10.21
- Jörgens, D., Smedby, Ö., and Moreno, R. (2018). Learning a single step of streamline tractography based on neural networks. In: *Computational Diffusion MRI: MICCAI Workshop, Québec, Canada, September 2017*. Berlin: Springer. p. 103–116.
- Kammen, A., Law, M., Tjan, B. S., Toga, A. W., and Shi, Y. (2016). Automated retinofugal visual pathway reconstruction with multi-shell HARDI and FOD-based analysis. *Neuroimage* 125, 767–779. doi: 10.1016/j.neuroimage.2015.11.005
- Karahan, E., Costigan, A. G., Graham, K. S., Lawrence, A. D., and Zhang, J. (2019). Cognitive and white-matter compartment models reveal selective relations between corticospinal tract microstructure and simple reaction time. *J. Neurosci.* 39, 5910–5921. doi: 10.1523/JNEUROSCI.2954-18.2019
- Karimi, D., Jaimes, C., Machado-Rivas, F., Vasung, L., Khan, S., Warfield, S. K., et al. (2021a). Deep learning-based parameter estimation in fetal diffusion-weighted MRI. *Neuroimage* 243, 118482. doi: 10.1016/j.neuroimage.2021.118482
- Karimi, D., Vasung, L., Jaimes, C., Machado-Rivas, F., Khan, S., Warfield, S. K., et al. (2021b). A machine learning-based method for estimating the number and orientations of major fascicles in diffusion-weighted magnetic resonance imaging. *Med. Image Anal.* 72, 102129. doi: 10.1016/j.media.2021.102129
- Karimi, D., Vasung, L., Jaimes, C., Machado-Rivas, F., Warfield, S. K., and Gholipour, A. (2021c). Learning to estimate the fiber orientation distribution function from diffusion-weighted MRI. *Neuroimage* 239, 118316. doi: 10.1016/j.neuroimage.2021.118316
- Kim, K.-H., Ronen, I., Formisano, E., Goebel, R., Ugurbil, K., and Kim, D.-S. (2004). Robust fiber tracking method by vector selection criterion in diffusion tensor images. In: *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Manhattan, NY: IEEE. p. 1080–1083.
- Klistorner, A., Vootakuru, N., Wang, C., Yiannikas, C., Graham, S. L., Parratt, J., et al. (2015). Decoding diffusivity in multiple sclerosis: analysis of optic radiation lesion and non-lesional white matter. *PLoS ONE* 10, e0122114. doi: 10.1371/journal.pone.0122114
- Koppers, S., Hauburger, C., and Merhof, D. (2017). Diffusion MRI signal augmentation: from single shell to multi shell with deep learning. In: *Computational Diffusion MRI: MICCAI Workshop, Athens, Greece, October 2016* 19. Berlin: Springer. p. 61–70.
- Kreher, B., Mader, I., and Kiselev, V. (2008). Gibbs tracking: a novel approach for the reconstruction of neuronal pathways. *Magn. Reson. Med.* 60, 953–963. doi: 10.1002/mrm.21749
- Lacerda, L. M., Clayden, J. D., Handley, S. E., Winston, G. P., Kaden, E., Tisdall, M., et al. (2020). Microstructural investigations of the visual pathways in pediatric epilepsy neurosurgery: insights from multi-shell diffusion magnetic resonance imaging. *Front. Neurosci.* 14, 269. doi: 10.3389/fnins.2020.00269
- Lazar, M., and Alexander, A. L. (2003). An error analysis of white matter tractography methods: synthetic diffusion tensor field simulations. *Neuroimage* 20, 1140–1153. doi: 10.1016/S1053-8119(03)00277-5
- Lazar, M., Weinstein, D. M., Tsuruda, J. S., Hasan, K. M., Arfanakis, K., Meyerand, M. E., et al. (2003). White matter tractography using diffusion tensor deflection. *Hum. Brain Mapp.* 18, 306–321. doi: 10.1002/hbm.10102
- Li, H., Liang, Z., Zhang, C., Liu, R., Li, J., Zhang, W., et al. (2021). SuperDTI: Ultrafast DTI and fiber tractography with deep learning. *Magn. Reson. Med.* 86, 3334–3347. doi: 10.1002/mrm.28937
- Li, S., He, J., Xue, T., Xie, G., Yao, S., Chen, Y., et al. (2022). DeepRGVP: A Novel Microstructure-Informed Supervised Contrastive Learning Framework for Automated Identification Of The Retinogeniculate Pathway Using dMRI Tractography. *arXiv [Preprint]*. doi: 10.48550/arXiv.2211.08119
- Li, M., He, J., and Feng, Y. (2020). Research progress of neural fiber tracking. *J. Image Graph.* 25, 1513–1528. doi: 10.11834/jig.190519
- Liu, M., Lerma-Usabiaga, G., Clascá, F., and Paz-Alonso, P. M. (2022). Reproducible protocol to obtain and measure first-order relay human thalamic white-matter tracts. *Neuroimage* 262, 119558. doi: 10.1016/j.neuroimage.2022.119558
- Liu, Y., and Gu, X. (2020). Evaluation and comparison of global-feature-based and local-feature-based segmentation algorithms in intracranial visual pathway delineation. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Manhattan, NY: IEEE. p. 1766–1769.
- Malcolm, J. G., Shenton, M. E., and Rathi, Y. (2010). Filtered multitensor tractography. *IEEE Trans. Med. Imaging* 29, 1664–1675. doi: 10.1109/TMI.2010.2048121
- Mansoor, A., Cerrolaza, J. J., Avery, R. A., and Linguraru, M. G. (2015). Partitioned shape modeling with on-the-fly sparse appearance learning for anterior visual pathway segmentation. In: *Clinical Image-Based Procedures. Translational Research in Medical Imaging: 4th International Workshop, CLIP 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015. Revised Selected Papers 4*. Berlin: Springer. p. 104–112.
- Mansoor, A., Cerrolaza, J. J., Idrees, R., Biggs, E., Alsharid, M. A., Avery, R. A., et al. (2016). Deep learning guided partitioned shape model for anterior visual pathway segmentation. *IEEE Trans. Med. Imaging* 35, 1856–1865. doi: 10.1109/TMI.2016.2535222
- Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* 34, 1993–2024. doi: 10.1109/TMI.2014.2377694
- Mlynarski, P., Delingette, H., Alghamdi, H., Bondiau, P.-Y., and Ayache, N. (2020). Anatomically consistent CNN-based segmentation of organs-at-risk in cranial radiotherapy. *J. Med. Imag.* 7, 014502–014502. doi: 10.1117/1.JMI.7.1.014502
- Mori, S., Crain, B. J., Chacko, V. P., and Van Zijl, P. C. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals Neurol.* 45, 265–269. doi: 10.1002/1531-8249(199902)45:2<265::aid-ana21>3.0.co;2-3
- Mutlu, U., Ikram, M. K., Roshchupkin, G. V., Bonnemaier, P. W. M., Colijn, J. M., Vingerling, J. R., et al. (2018). Thinner retinal layers are associated with changes in the visual pathway: A population-based study. *Hum. Brain Mapp.* 39, 4290–4301. doi: 10.1002/hbm.24246
- Neher, P. F., Côté, M.-A., Houde, J.-C., Descoteaux, M., and Maier-Hein, K. H. (2017). Fiber tractography using machine learning. *Neuroimage* 158, 417–429. doi: 10.1016/j.neuroimage.2017.07.028
- Noble, J. H., and Dawant, B. M. (2011). An atlas-navigated optimal medial axis and deformable model algorithm (NOMAD) for the segmentation of the optic nerves and chiasm in MR and CT images. *Med. Image Anal.* 15, 877–884. doi: 10.1016/j.media.2011.05.001
- Panda, S., Asman, A. J., Delisi, M. P., Mawn, L. A., Galloway, R. L., and Landman, B. A. (2014). Robust optic nerve segmentation on clinically acquired CT. In: *Medical Imaging 2014: Image Processing*. Bellingham, WA: SPIE. p. 362–371.
- Papadopoulou, A., Oertel, F. C., Chien, C., Kuchling, J., Zimmermann, H. G., Siebert, N., et al. (2021). Lateral geniculate nucleus volume changes after optic neuritis in neuromyelitis optica: a longitudinal study. *NeuroImage Clin.* 30, 102608. doi: 10.1016/j.nicl.2021.102608
- Parker, G. J., Haroon, H. A., and Wheeler-Kingshott, C. A. (2003). A framework for a streamline-based probabilistic index of connectivity (PICO) using a structural interpretation of MRI diffusion measurements. *J. Magn. Reson. Imag.* 18, 242–254. doi: 10.1002/jmri.10350
- Poulin, P., Côté, M.-A., Houde, J.-C., Petit, L., Neher, P. F., Maier-Hein, K. H., et al. (2017). Learn to track: deep learning for tractography. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11–13, 2017, Proceedings, Part I* 20. Berlin: Springer. p. 540–547.
- Poulin, P., Rheault, F., St-Onge, E., Jodoin, P.-M., and Descoteaux, M. (2018). Bundle-wise deep tracker: Learning to track bundle-specific streamline paths. Proc. of the Int. Society for Magnetic Resonance in medicine ISMRM-ESMRMB. Paris, France.
- Prins, D., Hanekamp, S., and Cornelissen, F. W. (2016). Structural brain MRI studies in eye diseases: are they clinically relevant? A review of current findings. *Acta Ophthalmol.* 94, 113–121. doi: 10.1111/aos.12825
- Puzniak, R. J., Mcpherson, B., Ahmadi, K., Herbig, A., Kaufmann, J., Liebe, T., et al. (2021a). CHIASM, the human brain albinism and achiasma MRI dataset. *Scientific data* 8, 308. doi: 10.1038/s41597-021-01080-w
- Puzniak, R. J., Prabhakaran, G. T., and Hoffmann, M. B. (2021b). Deep learning-based detection of malformed optic chiasm from MRI images. *Front. Neurosci.* 15, 1332. doi: 10.3389/fnins.2021.755785
- Ranjbarzadeh, R., Dorosti, S., Jafarzadeh Ghouschi, S., Safavi, S., Razmjoo, N., Tataei Sarshar, N., et al. (2022). Nerve optic segmentation in CT images using a deep learning model and a texture descriptor. *Comp. Intel. Syst.* 8, 3543–3557. doi: 10.1007/s40747-022-00694-w
- Raudaschl, P. F., Zaffino, P., Sharp, G. C., Spadea, M. F., Chen, A., Dawant, B. M., et al. (2017). Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. *Med. Phys.* 44, 2020–2036. doi: 10.1002/mp.12197
- Reich, D. S., Ozturk, A., Calabresi, P. A., and Mori, S. (2010). Automated vs. conventional tractography in multiple sclerosis: variability and correlation with disability. *Neuroimage* 49, 3047–3056. doi: 10.1016/j.neuroimage.2009.11.043
- Reid, L. B., Martinez-Heras, E., Manjon, J. V., Jeffree, R. L., Alexander, H., Trinder, J., et al. (2021). Fully automated delineation of the optic radiation for surgical planning using clinically feasible sequences. *Hum. Brain Mapp.* 42, 5911–5926. doi: 10.1002/hbm.25658
- Ren, X., Xiang, L., Nie, D., Shao, Y., Zhang, H., Shen, D., et al. (2018). Interleaved 3D-CNN s for joint segmentation of small-volume structures in head and neck CT images. *Med. Phys.* 45, 2063–2075. doi: 10.1002/mp.12837
- Romano, A., D'andrea, G., Minniti, G., Mastronardi, L., Ferrante, L., Fantozzi, L. M., et al. (2009). Pre-surgical planning and MR-tractography utility in brain tumour resection. *Eur. Radiol.* 19, 2798–2808. doi: 10.1007/s00330-009-1483-6

- Romano, A., Ferrante, M., Cipriani, V., Fasoli, F., Ferrante, L., D'andrea, G., et al. (2007). Role of magnetic resonance tractography in the preoperative planning and intraoperative assessment of patients with intra-axial brain tumours. *Radiol. Med.* 112, 906–920. doi: 10.1007/s11547-007-0181-1
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Berlin: Springer. p. 234–241.
- Rossi-Espagnet, M. C., Lucignani, M., Pasquini, L., Napolitano, A., Pro, S., Romano, A., et al. (2020). Visual pathways evaluation in Kearns Sayre syndrome: a diffusion tensor imaging study. *Neuroradiology* 62, 241–249. doi: 10.1007/s00234-019-02302-0
- Schurr, R., Duan, Y., Norcia, A. M., Ogawa, S., Yeatman, J. D., and Mezer, A. A. (2018). Tractography optimization using quantitative T1 mapping in the human optic radiation. *Neuroimage* 181, 645–658. doi: 10.1016/j.neuroimage.2018.06.060
- Sherbondy, A. J., Dougherty, R. F., Ben-Shachar, M., Napel, S., and Wandell, B. A. (2008). ConTrack: finding the most likely pathways between brain regions using diffusion tractography. *J. Vis.* 8, 15–15. doi: 10.1167/8.9.15
- Smith, M. M., and Strottmann, J. M. (2001). Imaging of the optic nerve and visual pathways. In: *Seminars in Ultrasound, CT and MRI*. Amsterdam: Elsevier. 473–487.
- Smith, R. E., Tournier, J.-D., Calamante, F., and Connelly, A. (2012). Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage* 62, 1924–1938. doi: 10.1016/j.neuroimage.2012.06.005
- Spierer, O., Ben Sira, L., Leibovitch, I., and Kesler, A. (2010). MRI demonstrates restricted diffusion in distal optic nerve in atypical optic neuritis. *J. Neuroophthalmol.* 30, 31–33. doi: 10.1097/WNO.0b013e3181ce18eb
- Takemura, M. Y., Hori, M., Yokoyama, K., Hamasaki, N., Suzuki, M., Kamagata, K., et al. (2017). Alterations of the optic pathway between unilateral and bilateral optic nerve damage in multiple sclerosis as revealed by the combined use of advanced diffusion kurtosis imaging and visual evoked potentials. *Magn. Reson. Imaging* 39, 24–30. doi: 10.1016/j.mri.2016.04.011
- Tamraz, J. C., Outin-Tamraz, C., and Saban, R. (1999). MR imaging anatomy of the optic pathways. *Radiol. Clin. North Am.* 37, 1–36. doi: 10.1016/S0033-8389(05)70076-2
- Tellouck, L., Durieux, M., Coupe, P., Cournard-Gregoire, A., Tellouck, J., Tourdias, T., et al. (2016). Optic radiations microstructural changes in glaucoma and association with severity: a study using 3Tesla-magnetic resonance diffusion tensor imaging. *Invest. Ophthalmol. Visual Sci.* 57, 6539–6547. doi: 10.1167/iov.16-19838
- Théberge, A., Desrosiers, C., Descoteaux, M., and Jodoin, P.-M. (2021). Track-to-Learn: A general framework for tractography with deep reinforcement learning. *Med. Image Anal.* 72, 102093. doi: 10.1016/j.media.2021.102093
- Tian, Q., Bilgic, B., Fan, Q., Liao, C., Ngamsombat, C., Hu, Y., et al. (2020). DeepDTI: High-fidelity six-direction diffusion tensor imaging using deep learning. *Neuroimage* 219, 117017. doi: 10.1016/j.neuroimage.2020.117017
- Tong, N., Gou, S., Yang, S., Cao, M., and Sheng, K. (2019). Shape constrained fully convolutional DenseNet with adversarial training for multiorgan segmentation on head and neck CT and low-field MR images. *Med. Phys.* 46, 2669–2682. doi: 10.1002/mp.13553
- Tong, N., Gou, S., Yang, S., Ruan, D., and Sheng, K. (2018). Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Med. Phys.* 45, 4558–4567. doi: 10.1002/mp.13147
- Tournier, J.-D., Calamante, F., Gadian, D. G., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage* 23, 1176–1185. doi: 10.1016/j.neuroimage.2004.07.037
- Tournier, J. D., Calamante, F., King, M., Gadian, D., and Connelly, A. (2002). Limitations and requirements of diffusion tensor fiber tracking: an assessment using simulations. *Magn. Reson. Med.* 47, 701–708. doi: 10.1002/mrm.10116
- Wang, C., Barton, J., Kyle, K., Ly, L., Barnett, Y., Hartung, H.-P., et al. (2021). Multiple sclerosis: structural and functional integrity of the visual system following alemtuzumab therapy. *Journal of Neurology Neurosurgery and Psychiatry* 92, 1319–1324. doi: 10.1136/jnnp-2021-326164
- Wang, L., Chitiboi, T., Meine, H., Günther, M., and Hahn, H. K. (2016). Principles and methods for automatic and semi-automatic tissue segmentation in MRI data. *MAGMA* 29, 95–110. doi: 10.1007/s10334-015-0520-5
- Wang, X., Li, Z., Huang, Y., and Jiao, Y. (2022). Multimodal medical image segmentation using multi-scale context-aware network. *Neurocomputing* 486, 135–146. doi: 10.1016/j.neucom.2021.11.017
- Wang, Y., Zhao, L., Wang, M., and Song, Z. (2019). Organ at risk segmentation in head and neck CT images using a two-stage segmentation framework based on 3D U-Net. *IEEE Access* 7, 144591–144602. doi: 10.1109/ACCESS.2019.2944958
- Wasserthal, J., Neher, P., and Maier-Hein, K. H. (2018). TractSeg-Fast and accurate white matter tract segmentation. *Neuroimage* 183, 239–253. doi: 10.1016/j.neuroimage.2018.07.070
- Wasserthal, J., Neher, P. F., Hirjak, D., and Maier-Hein, K. H. (2019). Combined tract segmentation and orientation mapping for bundle-specific tractography. *Med. Image Anal.* 58, 101559. doi: 10.1016/j.media.2019.101559
- Wegmayr, V., Giulieri, G., Holdener, S., and Buhmann, J. (2018). Data-driven fiber tractography with neural networks. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Manhattan, NY: IEEE. p. 1030–1033.
- Weinstein, D., Kindlmann, G., and Lundberg, E. (1999). Tensorlines: Advection-diffusion based propagation through diffusion tensor fields. In: *Proceedings Visualization '99 (Cat. No. 99CB37067)*. Manhattan, NY: IEEE. p. 249–250.
- Wichmann, W., and Müller-Forell, W. (2004). Anatomy of the visual system. *Eur. J. Radiol.* 49, 8–30. doi: 10.1016/j.ejrad.2003.11.001
- Winston, G. P., Daga, P., Stretton, J., Modat, M., Symms, M. R., Mcevoy, A. W., et al. (2012). Optic radiation tractography and vision in anterior temporal lobe resection. *Ann. Neurol.* 71, 334–341. doi: 10.1002/ana.22619
- Wu, C.-N., Duan, S.-F., Mu, X.-T., Wang, Y., Lan, P.-Y., Wang, X.-L., et al. (2019). Assessment of optic nerve and optic tract alterations in patients with orbital space-occupying lesions using probabilistic diffusion tractography. *Int. J. Ophthalmol.* 12, 1304. doi: 10.18240/ijo.2019.08.11
- Xu, H., Dong, M., Lee, M.-H., O'hara, N., Asano, E., and Jeong, J.-W. (2019). Objective detection of eloquent axonal pathways to minimize postoperative deficits in pediatric epilepsy surgery using diffusion tractography and convolutional neural networks. *IEEE Trans. Med. Imaging* 38, 1910–1922. doi: 10.1109/TMI.2019.2902073
- Xue, T., Zhang, F., Zhang, C., Chen, Y., Song, Y., Golby, A. J., et al. (2023). Superficial white matter analysis: An efficient point-cloud-based deep learning framework with supervised contrastive learning for consistent tractography parcellation across populations and dMRI acquisitions. *Med. Image Anal.* 85, 102759. doi: 10.1016/j.media.2023.102759
- Xue, T., Zhang, F., Zhang, C., Chen, Y., Song, Y., Makris, N., et al. (2022). Supwma: consistent and efficient tractography parcellation of superficial white matter with deep learning. In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI): Manhattan, NY: IEEE*. p. 1–5.
- Yamamoto, T., Yamada, K., Nishimura, T., and Kinoshita, S. (2005). Prognostic value of diffusion tensor imaging in acute optic neuritis. *Invest. Ophthalmol. Visual Sci.* 46, 639.
- Yang, X., Cerrolaza, J., Duan, C., Zhao, Q., Murnick, J., Safdar, N., et al. (2014). Weighted partitioned active shape model for optic pathway segmentation in MRI. In: *Clinical Image-Based Procedures. Translational Research in Medical Imaging: Third International Workshop, CLIP 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 14, 2014, Revised Selected Papers 3*. Berlin: Springer. p. 109–117.
- Yogarajah, M., Focke, N. K., Bonelli, S., Cercignani, M., Acheson, J., Parker, G., et al. (2009). Defining Meyer's loop-temporal lobe resections, visual field deficits and diffusion tensor tractography. *Brain* 132, 1656–1668. doi: 10.1093/brain/awp114
- Yoshimine, S., Ogawa, S., Horiguchi, H., Terao, M., Miyazaki, A., Matsumoto, K., et al. (2018). Age-related macular degeneration affects the optic radiation white matter projecting to locations of retinal damage. *Brain Struct. Funct.* 223, 3889–3900. doi: 10.1007/s00429-018-1702-5
- You, S., Lei, B., Wang, S., Chui, C. K., Cheung, A. C., Liu, Y., et al. (2022). Fine Perceptive GANs for Brain MR Image Super-Resolution in Wavelet Domain. Manhattan, NY: IEEE transactions on neural networks and learning systems.
- Zeng, R., Lv, J., Wang, H., Zhou, L., Barnett, M., Calamante, F., et al. (2022). FOD-Net: A deep learning method for fiber orientation distribution angular super resolution. *Med. Image Anal.* 79, 102431. doi: 10.1016/j.media.2022.102431
- Zhang, F., Hancock, E. R., Goodlett, C., and Gerig, G. (2009). Probabilistic white matter fiber tracking using particle filtering and von Mises–Fisher sampling. *Med. Image Anal.* 13, 5–18. doi: 10.1016/j.media.2008.05.001
- Zhang, F., Karayumak, S. C., Hoffmann, N., Rathi, Y., Golby, A. J., and O'donnell, L. J. (2020). Deep white matter analysis (DeepWMA): fast and consistent tractography segmentation. *Med. Image Anal.* 65, 101761. doi: 10.1016/j.media.2020.101761
- Zhao, Y., Tan, S., Chan, T. C. Y., Xu, Q., Zhao, J., Teng, D., et al. (2018). Clinical features of demyelinating optic neuritis with seropositive myelin oligodendrocyte glycoprotein antibody in Chinese patients. *Br. J. Ophthalmol.* 102, 1372–1377. doi: 10.1136/bjophthalmol-2017-311177
- Zhao, Z., Ai, D., Li, W., Fan, J., Song, H., Wang, Y., et al. (2019). Spatial probabilistic distribution map based 3D FCN for visual pathway segmentation. In: *Image and Graphics: 10th International Conference, ICIG 2019, Beijing, China, August 23–25, 2019, Proceedings, Part II 10*. Berlin: Springer. p. 509–518.
- Zhu, F., Gao, Z., Zhao, C., Zhu, Z., Tang, J., Liu, Y., et al. (2021). Semantic segmentation using deep learning to extract total extraocular muscles and optic nerve from orbital computed tomography images. *Optik* 244, 167551. doi: 10.1016/j.ijleo.2021.167551
- Zhu, W., Huang, Y., Zeng, L., Chen, X., Liu, Y., Qian, Z., et al. (2019). AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med. Phys.* 46, 576–589. doi: 10.1002/mp.13300
- Zolal, A., Sobottka, S. B., Podlessek, D., Linn, J., Rieger, B., Juratli, T. A., et al. (2016). Comparison of probabilistic and deterministic fiber tracking of cranial nerves. *J. Neurosurg.* 127, 613–621. doi: 10.3171/2016.8.JNS16363



OPEN ACCESS

EDITED BY

Zhengwang Wu,
University of North Carolina at Chapel Hill,
United States

REVIEWED BY

Hongjiang Wei,
Shanghai Jiao Tong University, China
Sahar Ahmad,
University of North Carolina at Chapel Hill,
United States

*CORRESPONDENCE

Yong Hu

✉ yhud@hku.hk

Shuqiang Wang

✉ sq.wang@siat.ac.cn

RECEIVED 10 April 2023

ACCEPTED 22 May 2023

PUBLISHED 13 June 2023

CITATION

Gong C, Jing C, Chen X, Pun CM, Huang G,
Saha A, Nieuwoudt M, Li H-X, Hu Y and Wang S
(2023) Generative AI for brain image computing
and brain network computing: a review.
Front. Neurosci. 17:1203104.
doi: 10.3389/fnins.2023.1203104

COPYRIGHT

© 2023 Gong, Jing, Chen, Pun, Huang, Saha,
Nieuwoudt, Li, Hu 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.

Generative AI for brain image computing and brain network computing: a review

Changwei Gong^{1,2}, Changhong Jing^{1,2}, Xuhang Chen^{1,3},
Chi Man Pun³, Guoli Huang¹, Ashirbani Saha⁴, Martin Nieuwoudt⁵,
Han-Xiong Li⁶, Yong Hu^{7*} and Shuqiang Wang^{1,2*}

¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China,

²Department of Computer Science, University of Chinese Academy of Sciences, Beijing, China,

³Department of Computer and Information Science, University of Macau, Macau, China, ⁴Department of Oncology and School of Biomedical Engineering, McMaster University, Hamilton, ON, Canada, ⁵Institute for Biomedical Engineering, Stellenbosch University, Stellenbosch, South Africa, ⁶Department of Systems Engineering, City University of Hong Kong, Hong Kong, China, ⁷Department of Orthopaedics and Traumatology, The University of Hong Kong, Hong Kong, China

Recent years have witnessed a significant advancement in brain imaging techniques that offer a non-invasive approach to mapping the structure and function of the brain. Concurrently, generative artificial intelligence (AI) has experienced substantial growth, involving using existing data to create new content with a similar underlying pattern to real-world data. The integration of these two domains, generative AI in neuroimaging, presents a promising avenue for exploring various fields of brain imaging and brain network computing, particularly in the areas of extracting spatiotemporal brain features and reconstructing the topological connectivity of brain networks. Therefore, this study reviewed the advanced models, tasks, challenges, and prospects of brain imaging and brain network computing techniques and intends to provide a comprehensive picture of current generative AI techniques in brain imaging. This review is focused on novel methodological approaches and applications of related new methods. It discussed fundamental theories and algorithms of four classic generative models and provided a systematic survey and categorization of tasks, including co-registration, super-resolution, enhancement, classification, segmentation, cross-modality, brain network analysis, and brain decoding. This paper also highlighted the challenges and future directions of the latest work with the expectation that future research can be beneficial.

KEYWORDS

generative models, brain imaging, brain network, diffusion model, generative adversarial network, variational autoencoder

1. Introduction

Brain imaging, providing a way to non-invasively map the structure and function of the brain, has developed significantly in recent years (Gui et al., 2010). For instance, functional brain imaging, such as functional magnetic resonance imaging (fMRI), has the potential to revolutionize researchers' understanding of the physical basis of the brain and offers a powerful tool to understand how the brain adapts to various cognitive activities and tasks (Allen et al., 2014). Additionally, it offers a powerful tool that assists in understanding how the brain adapts to various cognitive activities and tasks. Generative artificial intelligence refers to new technologies that employ existing data including images, text, and audio files

to create new content. This new content has a similar underlying pattern of real-world data and has great potential applications in many areas. Synthetic data from generative AI (Wang et al., 2017; Lei et al., 2020b) can train machine learning models (Liu Y. et al., 2021; Lei et al., 2022b) to be less biased and help robots to learn more abstract concepts both in the real and virtual world. The development of neuroimaging as a cross-discipline between imaging and neuroscience has enabled the qualitative and quantitative analysis of images in multiple dimensions. Neuroimaging is a powerful tool for studying brain science, revealing the anatomical structure and working mechanisms of the brain, as well as diagnosing and treating brain diseases. The synergistic developments between emerging analytic technologies and data-sharing initiatives have the potential to transform the role of neuroimaging in clinical applications. While basic neuroscience focuses on understanding how brain activity produces behavior, clinical applications aim to develop tools that are useful for clinical decision-making and treatment development.

Brain images reveal multiple modalities due to different imaging principles and techniques. As shown in Figure 1, multi-modality brain imaging contains many different types, such as diffusion tensor imaging (DTI), fluid-attenuated inversion recovery (FLAIR) MRI, Susceptibility weighted imaging (SWI) MRI, resting state functional MRI (rs-fMRI) and fluorodeoxyglucose positron emission tomography (FDG-PET), etc. Brain imaging can be divided into two broad categories such as functional and structural imaging. Functional neuroimaging, which has generated great optimism about its potential to both revolutionize researchers' understanding of the physical basis of the brain and to provide clinically useful tools (Yu et al., 2021b), has made significant progress in achieving the former goal. However, functional neuroimaging results and models have yet to be incorporated into clinical practice. For decades, numerous translational neuroimaging and radiological studies have identified the characteristics that predict health-related outcomes (Wang S.-Q. et al., 2015; Wang et al., 2020a), including current diagnostic categories and measures of symptoms (Lei et al., 2022a), cognitive and affective computing processes, and cognitive performance. Redefining diagnostic categories, identifying neuropathological features, and assessing healthy brain function outside of current clinical diagnostic categories are potential outcomes of such studies.

Further analysis of brain images can provide morphological information about brain regions, such as their volume, thickness, and surface area. Automated computer analysis has replaced expert anatomists' manual labeling of brain images. In voxel-based morphometry, voxels are segmented into one of three tissue categories (cerebrospinal fluid, white matter, or gray matter) based on their image intensity. After recording all scans in the study into a common anatomical space, the gray matter density of each voxel can be compared between the whole brain and the subject, using the average brain as a template. This process extracts graphical data information about brain-related patterns from brain imaging voxels with high-resolution structures. Lundervold and Lundervold (2019) demonstrated that the introduction of automated computer analysis of magnetic resonance imaging has facilitated the *in vivo* study of whole-brain coordinated patterns in thousands of individuals (Hu et al., 2020b).

Moreover, in the field of network neuroscience (Bassett and Sporns, 2017), the theory and applications of generative AI offer

a powerful tool for brain imaging and brain network computing including but not limited to extraction of brain spatiotemporal features and the reconstruction of the topological connectivity of brain networks (Calhoun et al., 2014; Gong et al., 2022). Brain networks, which represent the global connectivity of the brain's structure and function, are crucial in understanding the neural basis of cognitive processes, neuroanatomy, functional brain imaging, and neurodevelopment. Brain network computing involves the construction, reconstruction, analysis, and optimization of brain networks. While brain imaging allows for the qualitative and quantitative analysis of the brain's anatomical and functional structure in two or three dimensions, brain network computing enables the study of brain topological features and covariant features (Isallari and Rekik, 2021). Various tools, such as PANDA (Cui et al., 2013) and GREYNA (Wang J. et al., 2015), can be used for constructing brain networks. However, the brain networks produced by these tools are subjective, time-consuming, and depend on the operator's experience. This review also surveys the development of AI-based algorithms that can automatically construct brain networks.

This paper provides a brief overview of research related to generative learning models in brain imaging from three different perspectives: AI-based generative models, tasks for brain imaging, and the prospects of generative AI for brain imaging. The paper reviews recent developments and advancements made in each of these areas.

2. Generative learning model

The generative learning model refers to a class of machine learning (ML) models that can generate new data similar to the training data on which they were trained. Large-scale generative models are trained on massive datasets and require specialized hardware, such as GPUs, for efficient training. As shown in Figure 2, several types of generative models exist, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Flow Models, and Denoising Diffusion Probabilistic Models (DDPMs). Introduced by Goodfellow et al. (2014), GANs are a type of neural network comprising two parts: a generator network that creates new data and a discriminator network that distinguishes between real and fake data. VAEs, proposed by Kingma and Welling (2013), generate new data by learning a compressed representation of the input data. Flow Models, proposed by Rezende and Mohamed (2015), model the probability distribution of the input data and invert it. DDPMs, a new type of generative model introduced by Ho et al. (2020), have gained popularity in recent years. They draw inspiration from the physical process of gas molecule diffusion, in which molecules diffuse from high-density to low-density areas. DDPMs learn to model the data distribution from input data incrementally.

2.1. Variational autoencoder

Autoencoders are a type of neural network that encode the input X into a low-dimensional vector z , also known as the latent space, and then reconstruct the input X based on z . By minimizing the error between X and the generated output \hat{X} , autoencoders

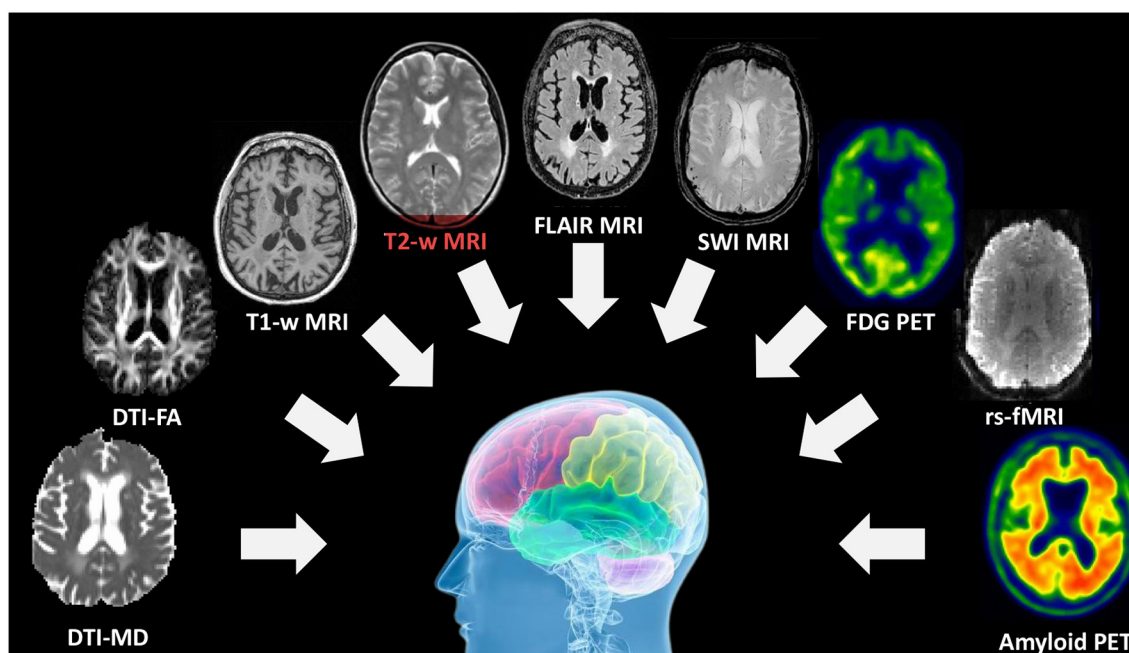


FIGURE 1
Multi-modality brain imaging includes DTI, sMRI, fMRI, PET, and other imaging types.

are trained to gradually reduce the reconstruction error, thereby achieving the goal of reconstruction. However, autoencoders suffer from the limitation of not being able to generate new content as they cannot produce latent vectors arbitrarily. This is because the latent vectors z are all encoded from the original images by the encoder.

In Equation (1), where a set of raw data samples X_1, \dots, X_n is available to describe the population, direct sampling from the probability distribution $p(X)$ would be feasible if $p(X)$ were known. However, in practice, the distribution of the raw data $p(X)$ is typically unknown.

$$p(X) = \sum_Z p(X|Z)p(Z) \quad (1)$$

$$\begin{aligned} p(Z) &= \sum_X p(Z|X)p(X) \\ &= \mathcal{N}(0, I) \sum_X p(X) \\ &= \mathcal{N}(0, I) \end{aligned} \quad (2)$$

To address this issue, researchers have added constraints to the latent space Z (the space corresponding to the latent vectors) to impose a prior distribution on the latent vectors. This led to the development of the variational autoencoder (VAE) model, which adds a constraint to the encoder to force it to produce latent variables that follow a normal distribution in Equation (2). It is this constraint that distinguishes VAE from traditional autoencoders.

A key aspect of variational autoencoders (VAEs) is the addition of a constraint that enforces a normal distribution on the latent space Z . Determining this normal distribution is the primary

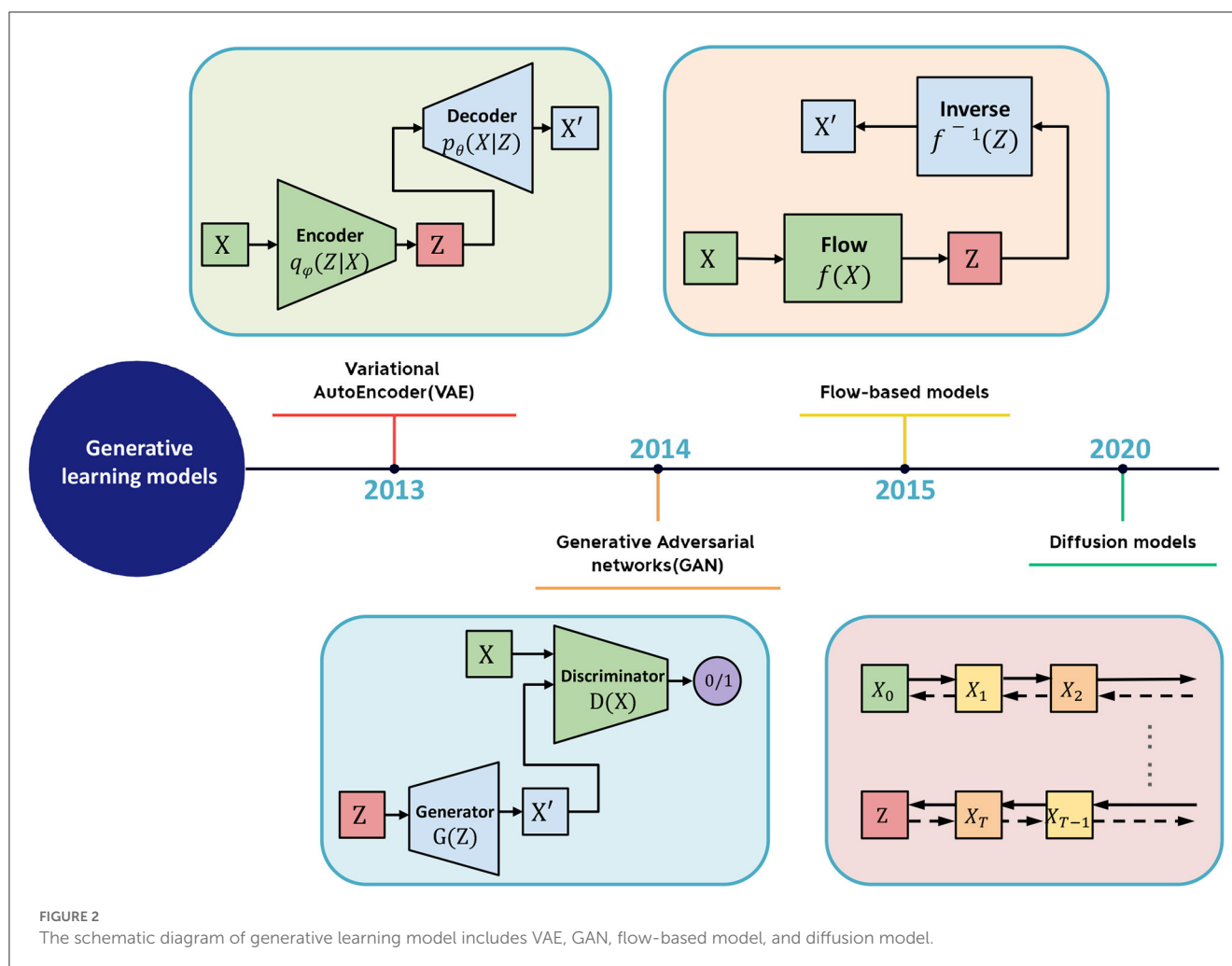
objective of VAEs. To specify a normal distribution, two parameters must be determined: the mean μ and the standard deviation σ . To accomplish this, the encoder encodes input samples X into two latent dimension vectors, μ and σ , which represent the mean and variance of the latent space assumed to follow a normal distribution (Mo and Wang, 2009). To sample Z from this latent space, VAE assumes that the latent normal distribution can generate the input images. VAE first samples a random vector ϵ from the standard normal distribution $N(0, I)$, and then computes:

$$Z = \mu + \sigma \odot \epsilon \quad (3)$$

Here \odot denotes element-wise multiplication. Z is a vector sampled from the latent space, and Z is used as input to the decoder to generate a reconstructed sample \hat{X} . The above steps constitute the forward propagation process of the entire network. To perform backpropagation, the loss function is evaluated from two aspects: the similarity between the generated output \hat{X} and the original input X and the similarity between the distribution of the latent space and the normal distribution. The similarity between X and \hat{X} is generally measured using reconstructions loss, while the similarity between two distributions is generally measured using the Kullback-Leibler divergence (Joyce, 2011).

$$KL(p(x)||q(x)) = \int p(x) \ln \frac{p(x)}{q(x)} dx \quad (4)$$

$$\begin{aligned} Loss(X, \hat{X}) &= Loss_{reconstruction} \\ &+ \beta Loss_{KL}(z, \mathcal{N}(0, I_d)) \end{aligned} \quad (5)$$



In summary, VAEs have found useful applications in brain imaging. VAEs can effectively cluster similar patterns in brain activity data and detect subtle changes that may not be easily perceptible to the human eye (Tezcan et al., 2017; Cheng et al., 2021). Furthermore, the learned lower-dimensional representation by VAEs can also serve as a data compression method to minimize computational resources when analyzing complex brain network data (Qiao et al., 2021). The ability to model complex data and generate new data points that resemble the original ones makes VAEs a powerful tool in gaining insights into the underlying mechanisms of neurological disorders and diseases (Zhao et al., 2019).

2.2. Generative adversarial network

A Generative Adversarial Network (GAN) is a type of machine learning framework designed by Goodfellow et al. (2014). GANs are composed of two neural networks that compete against each other in a zero-sum game, where one agent's gain is another agent's loss. The framework learns to generate new data with the same statistics as the training set, enabling the GAN to generate new data that

resembles the original data.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (6)$$

GANs are based on game theory and can be viewed as a two-player minimax game, where the generator aims to minimize the difference between the distribution of the generated samples and the distribution of the real data, while the discriminator aims to maximize the difference between the two distributions. During training, the generator tries to produce samples that can mislead the discriminator into thinking they are real, while the discriminator tries to correctly distinguish between real and synthetic samples.

The generator usually consists of a series of deconvolutional layers that gradually upsample the random input vector into a sample that is intended to resemble the training data. The discriminator usually consists of a series of convolutional layers that downsample the input image or sample into a lower-dimensional feature representation, followed by a few fully connected layers that compute the final prediction. The loss function used in GANs is typically the binary cross-entropy loss, which measures the difference between the predicted

probabilities of the discriminator and the true labels (0 for synthetic samples and 1 for real samples). Other loss functions such as Wasserstein distance or hinge loss have been proposed to address some of the limitations of the binary cross-entropy loss. However, the original GAN suffers from the issue of gradient vanishing, which can lead to unstable training and poor sample quality. One of the main challenges of training GANs is the mode collapse problem, where the generator produces a limited variety of samples that are similar to each other, rather than generating diverse samples that cover the entire range of the training data. Several techniques have been proposed to overcome this problem, such as adding noise to the input of the discriminator, using different types of regularization, or using multi-scale or multi-modal architectures. To overcome this limitation, methods such as WGAN were introduced, which use the Wasserstein distance to measure the distance between the real and generated distributions.

$$W_c[\tilde{p}(x), q(x)] = \inf_{\gamma \in \Pi(\tilde{p}(x), q(x))} \mathbb{E}_{(x,y) \sim \gamma} [c(x,y)] \quad (7)$$

The Wasserstein GAN (WGAN), proposed by Arjovsky et al. (2017), aims to overcome the limitations of the original GAN model by using the Wasserstein distance to measure the distance between the real and generated distributions. The objective of WGAN is to minimize the optimal transport cost function, which represents the minimum cost of transforming the generated distribution $q(x)$ into the real distribution $\tilde{p}(x)$ through a series of small steps. The cost of each step is measured by the cost function $c(x, y)$, which represents the distance between the samples x and y . By using the Wasserstein distance instead of the Jensen-Shannon divergence used in the original GAN, WGAN is able to provide more stable training and generate higher quality samples.

$$\arg \min_G \arg \max_{T, \|T\|_L \leq 1} \mathbb{E}_{x \sim \tilde{p}(x)} [T(x)] - \mathbb{E}_{x \sim q(z)} [T(G(z))] \quad (8)$$

WGAN, which uses the Wasserstein distance instead of the Jensen-Shannon divergence used in the original GAN, provides more stable training and generates higher quality samples. WGAN also has other advantages, such as improved convergence properties and the ability to measure the distance between distributions more accurately. Overall, WGAN represents a significant advancement in the field of generative modeling and has been successfully applied in various applications, such as image generation, data augmentation, and domain adaptation. Its success has led to the development of several variants, such as Wasserstein GAN with Gradient Penalty (WGAN-GP; Gulrajani et al., 2017), which further improves the stability and efficiency of training.

Another widely used GAN variant in the medical field is CycleGAN, a type of unsupervised learning technique proposed by Zhu et al. (2017), which can learn the mapping between two different domains without any paired data. CycleGAN has several advantages, such as its ability to learn the mapping between two domains without the need for paired data and its ability to handle multimodal and many-to-many mappings. It has been successfully applied in various applications, including

medical image analysis, such as image-to-image translation, segmentation, and registration. CycleGAN has also inspired the development of several variants, such as DualGAN (Yi et al., 2017), DiscoGAN (Kim et al., 2017), and UNIT (Liu et al., 2017), which further improve the performance and versatility of the original CycleGAN. The main idea behind CycleGAN is to use two generators and two discriminators to learn the mapping between the domains. The formula for CycleGAN is as follows:

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y) \quad (9)$$

The two generators in CycleGAN are used to generate images from one domain and then transform them into images from the other domain. The two discriminators are used to distinguish between the generated images and the real images from the other domain. The CycleGAN objective function includes two GAN losses, which encourage the generators to generate realistic images, and a cycle-consistency loss, which encourages the generators to learn a mapping between the two domains.

$$\begin{aligned} \mathcal{L}(G, F, D_X, D_Y) = & \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ & + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ & + \lambda \mathcal{L}_{\text{cyc}}(G, F) \end{aligned} \quad (10)$$

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = & \mathbb{E}_{y \sim p_{\text{tan}}(y)} [\log D_Y(y)] \\ & + \mathbb{E}_{x \sim p_{\text{tan}}(x)} [\log(1 - D_Y(G(x)))] \end{aligned} \quad (11)$$

$$\begin{aligned} \mathcal{L}_{\text{cyc}}(G, F) = & \mathbb{E}_{x \sim p_{\text{tan}}(x)} [\|F(G(x)) - x\|_1] \\ & + \mathbb{E}_{y \sim p_{\text{tan}}(y)} [\|G(F(y)) - y\|_1] \end{aligned} \quad (12)$$

Overall, GANs have found useful applications in brain imaging and network analysis. They can generate synthetic data samples that resemble real data (Dar et al., 2019), enabling researchers to explore brain activity patterns and identify underlying structures. GANs can also augment data by generating synthetic samples to balance imbalanced classes in the dataset, improving deep learning performance in tasks such as image segmentation and classification (Gao et al., 2021). Also, GANs (Lei et al., 2020a) can generate new brain activity patterns (Zuo et al., 2021) in brain network analysis, which can be used to simulate brain activity under various conditions and understand how the network responds to different stimuli. GANs (Wang et al., 2020b) can also help model the relationships between different brain regions and predict the functional connectivity patterns of the brain.

2.3. Flow-based generative model

Flow-based generative models are a type of deep generative model that can learn to generate new samples similar to a given dataset. Flow-based models are based on the concept of normalizing flows, which are transformations that can map a

simple distribution (e.g., Gaussian) to a more complex distribution (e.g., the distribution of the training data). Flow-based models have been applied to a wide range of applications, such as image generation, video generation, text generation, and even molecular design. Several variations of flow-based models have been proposed, such as conditional flow-based models, which can generate samples conditioned on a given input, and autoregressive flow-based models, which can generate samples by sequentially generating each dimension of the sample.

Flow-based models consist of a series of invertible transformations that map a simple distribution to the distribution of the training data. The inverse of each transformation is also computable, which allows for efficient computation of the likelihood of the data and generation of new samples. The transformations can be learned using maximum likelihood estimation or other methods.

$$G^* = \operatorname{argmax}_G \sum_{i=1}^m \log P_G(x^i) \quad (13)$$

$$\approx \operatorname{argmin}_{G \sim \text{KL}(P_{\text{data}} \| P_G)}$$

$$\log q(\mathbf{x}) = -\frac{D}{2} \log(2\pi) - \frac{1}{2} \|\mathbf{f}(\mathbf{x})\|^2 + \log \left| \det \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right] \right| \quad (14)$$

During training, the flow-based model learns to maximize the likelihood of the training data, which is typically computed using the change of variables formula and the likelihood of the simple distribution. The model can be trained using stochastic gradient descent or other optimization methods. It has several advantages over other types of generative models, such as explicit likelihood computation, efficient sampling, and the ability to perform exact inference. However, they also have some limitations, such as the requirement of invertible transformations, which can restrict the expressiveness of the model.

In summary, Flow-Based Generative Models offer a promising approach for modeling complex data distributions and have potential applications in brain imaging and brain network research. These models can accurately cluster brain activity patterns, identify the structure of the data, and generate synthetic data that resemble the real samples (Dong et al., 2022). Additionally, Flow-Based Models can be used to learn a lower-dimensional representation of the functional connectivity patterns in the brain, enabling researchers to identify relevant features for predicting network changes. The direct modeling of likelihood and the ability to generate novel samples make these models a powerful tool in understanding the underlying mechanisms of complex systems.

2.4. Diffusion model

Diffusion models belong to the category of latent variable models in machine learning that utilize Markov chains and variational inference to discern the underlying structure of a dataset. They offer a promising avenue for deep generative modeling owing to their straightforward training process, robust expressive capacity, and ability to generate

data via ancestral sampling without the prerequisite of a posterior distribution.

$$\begin{aligned} \log p(\mathbf{x}) &\geq \mathbb{E}_{q(\mathbf{x}_{1:T} | \mathbf{x}_0)} \left[\log \frac{p(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T} | \mathbf{x}_0)} \right] \\ &= \underbrace{\mathbb{E}_{q(\mathbf{x}_1 | \mathbf{x}_0)} [\log p_\theta(\mathbf{x}_0 | \mathbf{x}_1)]}_{\text{reconstruction term}} \\ &\quad - \underbrace{\mathbb{E}_{q(\mathbf{x}_{T-1} | \mathbf{x}_0)} [D_{\text{KL}}(q(\mathbf{x}_T | \mathbf{x}_{T-1}) \| p(\mathbf{x}_T))]}_{\text{prior matching term}} \\ &\quad - \sum_{t=2}^T \underbrace{\mathbb{E}_{q(\mathbf{x}_t | \mathbf{x}_0)} [q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)]}_{\text{denoising matching term}} \end{aligned} \quad (15)$$

The optimization of the diffusion model culminates in training a neural network to predict the original image from any time step of the noise image as input, with the optimization objective being to minimize the prediction error. Moreover, the optimization of the noise-matching term in equation 15 can be approximated by minimizing the expected prediction error at each time step using random sampling.

$$L_0 = -\log p_\theta(\mathbf{x}_0 | \mathbf{x}_1) \quad (16)$$

$$L_{t-1} = D_{\text{KL}}(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \| p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t))$$

$$L_T = D_{\text{KL}}(q(\mathbf{x}_T | \mathbf{x}_0) \| p(\mathbf{x}_T))$$

$$\arg \min_{\theta} \mathbb{E}_{t \sim U\{2, T\}} [\mathbb{E}_{q(\mathbf{x}_t | \mathbf{x}_0)} [D_{\text{KL}}(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0))]] \quad (17)$$

In contrast to other deep generative models such as VAE, GAN, and normalizing flow, diffusion models offer unique advantages while overcoming several limitations and challenges. The training of VAEs can be challenging due to the difficulty in selecting the variational posterior, while GANs require an additional discriminator network, and normalizing flow models have limited expressive power. In contrast, diffusion models utilize the diffusion process of data points through the latent space to derive a solution that involves training only a generator with a simple objective function, without the need for training other networks.

In computer vision, diffusion models train neural networks to denoise images blurred with Gaussian noise by learning to reverse the diffusion process. Three examples of generic diffusion modeling frameworks used in computer vision include denoising diffusion probabilistic models, noise-conditioned score networks, and stochastic differential equations. In brain imaging and brain network analysis, diffusion models serve as a valuable tool for estimating the underlying structure of brain function and structure (Chung and Ye, 2022), which is essential for understanding the mechanisms of neurological disorders and diseases (Myronenko,

2018). By modeling the diffusion of data points through the latent space, diffusion models are capable of effectively capturing changes in brain connectivity over time and identifying regions critical to brain structure (Wolleb et al., 2022). Additionally, diffusion models can simulate brain activity under different conditions and predict how the brain network will respond to various stimuli.

Overall, diffusion models provide a valuable approach to modeling latent spaces in various fields, including computer vision, brain imaging, and brain network analysis. By offering an efficient and effective means of estimating the underlying structure of datasets, diffusion models can be a powerful tool for gaining insights into the spatiotemporal dynamics of large and complex systems.

3. Tasks for brain imaging and brain network construction

In this section, tasks in brain imaging and brain network construction are specifically categorized and investigated in eight categories, including co-registration and super-resolution (shown in Figure 3), enhancement and classification (shown in Figure 4), segmentation and cross-modality (shown in Figure 5), and brain network analysis and brain decode (shown in Figure 6).

3.1. Co-registration

Co-registration is a crucial step in medical image analysis to align images from different modalities or time points. However, it is a challenging task due to various factors, such as noise, artifacts, motion, and anatomical differences. Many innovative methods have been proposed to tackle these challenges and improve co-registration performance.

For instance, Sundar et al. (2021) proposed conditional GANs to address intra-frame motion problems in dynamic PET studies of the brain. Yang et al. (2020) proposed indirect multimodal image registration and completion by using synthetic CT images obtained from multi-contrast MRI. Kong et al. (2021) introduced RegGAN for image-to-image translation and registration which includes noise reduction. Furthermore, Wang B. et al. (2022) proposed invertible AC-flow for direct generation of attenuation-corrected PET images without CT or MR images. Apart from these deep learning-based methods, a diffusion-based image registration method called DiffuseMorph was introduced by Kim et al. (2022). This method overcomes the limitations of traditional and deep learning-based methods due to computational complexity and topological folding.

These proposed methods have shown promising results in improving co-registration performance in medical imaging. Future research can explore further advances to overcome the remaining challenges, such as reducing the time taken for co-registration while maintaining high accuracy and improving the robustness and generalization of existing solutions.

3.2. Super-resolution

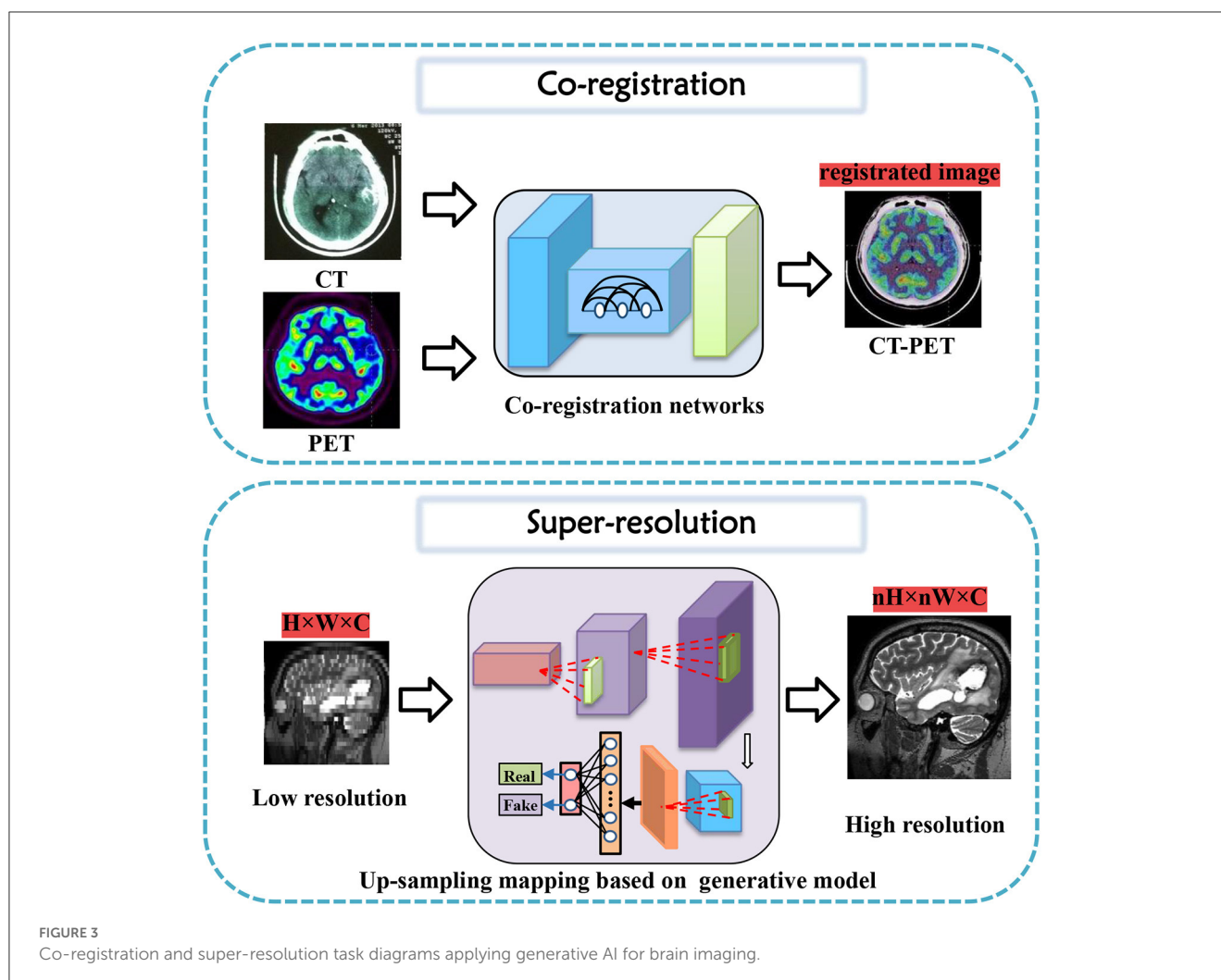
Research into high-resolution brain imaging has yielded promising results, with generative models proving to be a popular and effective approach (Sun et al., 2022). One such approach, as described in Song et al. (2020), involves using a GAN architecture with anatomical and spatial inputs for creating super-resolved brain PET images. According to the authors, the proposed GAN outperforms other deep learning models and penalized deconvolution techniques. Similarly, You et al. (2022) suggests using fine perceptive generative adversarial networks (FP-GANs) for high-resolution magnetic resonance imaging. This technique applies a sub-band generative adversarial network and sub-band attention for super-resolution in individual sub-bands.

These studies contribute to the growing body of literature on super-resolution tasks for brain imaging, with generative models emerging as promising solutions. The success of these models suggests that they could be applied to other high-resolution imaging tasks requiring greater detail and precision (Wicaksono et al., 2022). However, further research is needed to fully evaluate the performance and potential limitations of generative models for these tasks.

3.3. Enhancement

Data enhancement is a widely adopted approach in improving the performance of deep learning models for various medical image analysis tasks. Some studies cater to the task of data enhancement in medical tasks using generative AI. The first CycleGAN-based method for MR-to-CT synthesis proposed by Wolterink et al. (2017), shows that this technique can generate high-quality synthetic CT scans that are similar in appearance to real ones. Both the designed GAN model and novel loss function take enhancement tasks a step further in their performance (Dar et al., 2019).

Similarly, Yurt et al. (2021) proposed a multi-stream approach that integrates multiple source images to synthesize missing multi-contrast MRI images, outperforming other state-of-the-art methods. Zhan et al. (2021) proposed a Multi-scale Gate Mergence based GAN model that accurately diagnoses patients with corrupted image sequences by weighing different modalities across locations. Luo et al. (2021) proposed an edge-preserving MRI image synthesis GAN model, infusing an auxiliary edge image generation task to help preserve edge information and improve latent representation features, and an iterative multi-scale fusion module to further improve the quality of the synthesized target modality. Recently, Upadhyay et al. (2021) proposed a robust GAN-based framework that models an adaptive loss function to improve robustness to out-of-distribution (OOD)-noisy data and estimates per-voxel uncertainty in predictions for image-to-image translation across two real-world datasets in medical imaging applications. Luo et al. (2022) proposed an adaptive rectification-based GAN model with spectral constraint to synthesize high-quality standard-dose PET images from low-dose PET images, reducing radiation exposure while maintaining accurate diagnoses.



These studies demonstrate the promising potential of deep learning-based approaches and data enhancement in enhancing the quality and performance of medical image generation tasks.

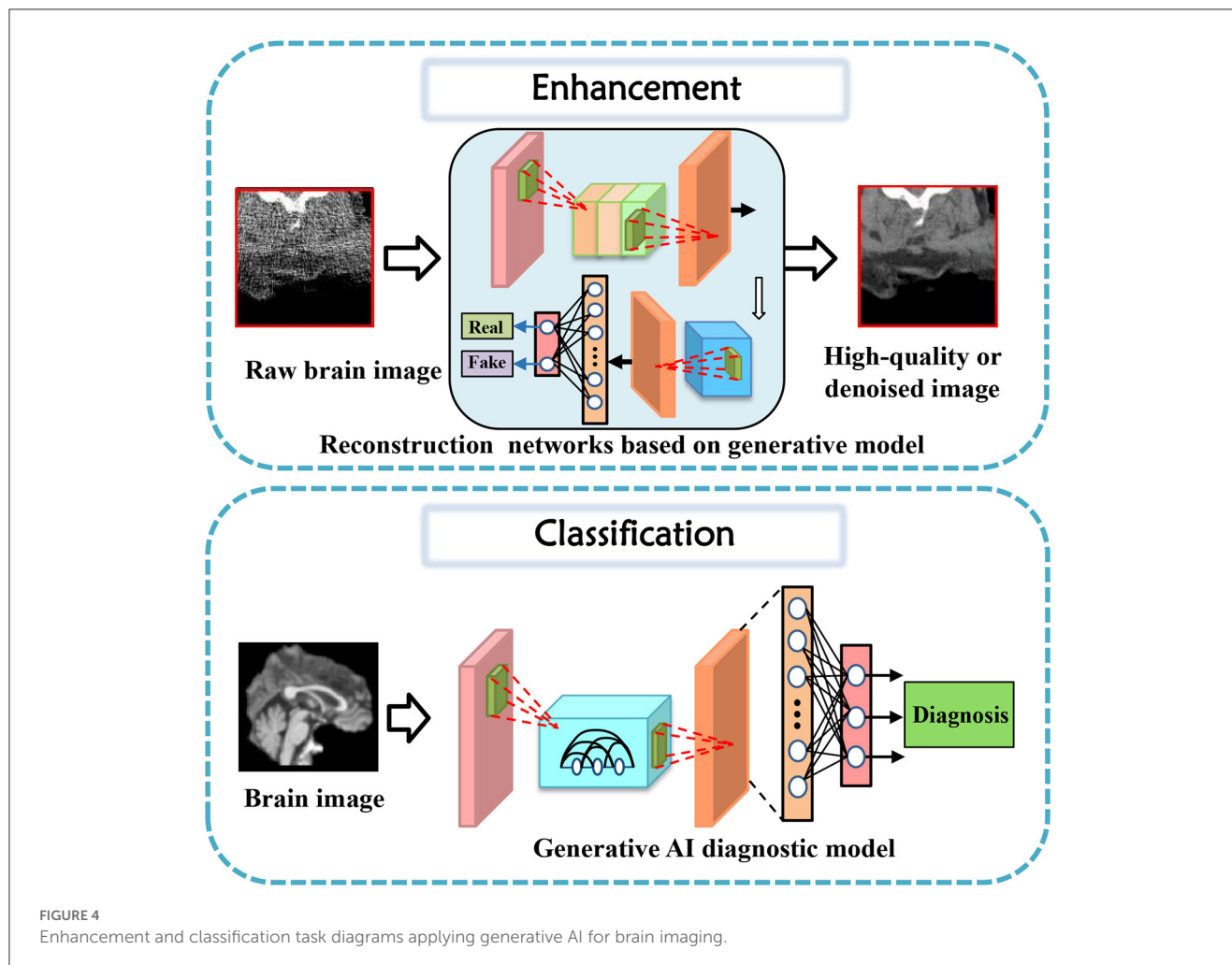
3.4. Classification

Classification of brain diseases is a crucial task for early diagnosis and effective treatment. The advancements in deep learning techniques have led to the development of various generative models for the automatic classification of neuroimages. Recently, several generative models were proposed to focus on brain disease classification tasks.

In 2019, Pan et al. (2019) propose a unified deep learning framework to jointly perform image synthesis and disease diagnosis using incomplete multi-modal neuroimaging data. The proposed method includes two networks: a Disease-Image Specific Neural Network (DSNN) to capture the spatial information of MRI/PET scans and a Feature-consistent Generative Adversarial Network (FGAN) to synthesize missing images by encouraging DSNN feature maps of synthetic images and their respective real images to be consistent. The method achieves state-of-the-art performance for Alzheimer's disease identification and mild

cognitive impairment conversion prediction tasks. Besides, pattern expression offered complementary performance to biomarkers in predicting clinical progression, making these deep-learning-derived biomarkers promising tools for precision diagnostics and targeted clinical trial recruitment. Yang Z. et al. (2021) applied deep learning framework to longitudinal data and revealed two distinct progression pathways that were predictive of future neurodegeneration rates.

Indeed, there are several generative models that have been designed with a deeper consideration of prior settings for tasks such as biomarkers and clinical reports. Wang et al. (2020a) propose an ensemble of 3D convolutional neural networks (CNNs) with dense connections for automatic diagnosis of Alzheimer's disease (AD) and mild cognitive impairment (MCI). The proposed model was evaluated on the ADNI dataset using a probability-based fusion method that combines multiple architectures. Shin et al. (2020) propose a GAN-based approach for the diagnosis of Alzheimer's Disease (AD) using T1-weighted MRIs as input data. The authors incorporate AD diagnosis into the training objective to achieve better classification performance. This architecture shows state-of-the-art results for three- or four-class classification tasks involving MCI, normal cognition, or Alzheimer's disease. Kim et al. (2020) propose a GAN-based model for classifying



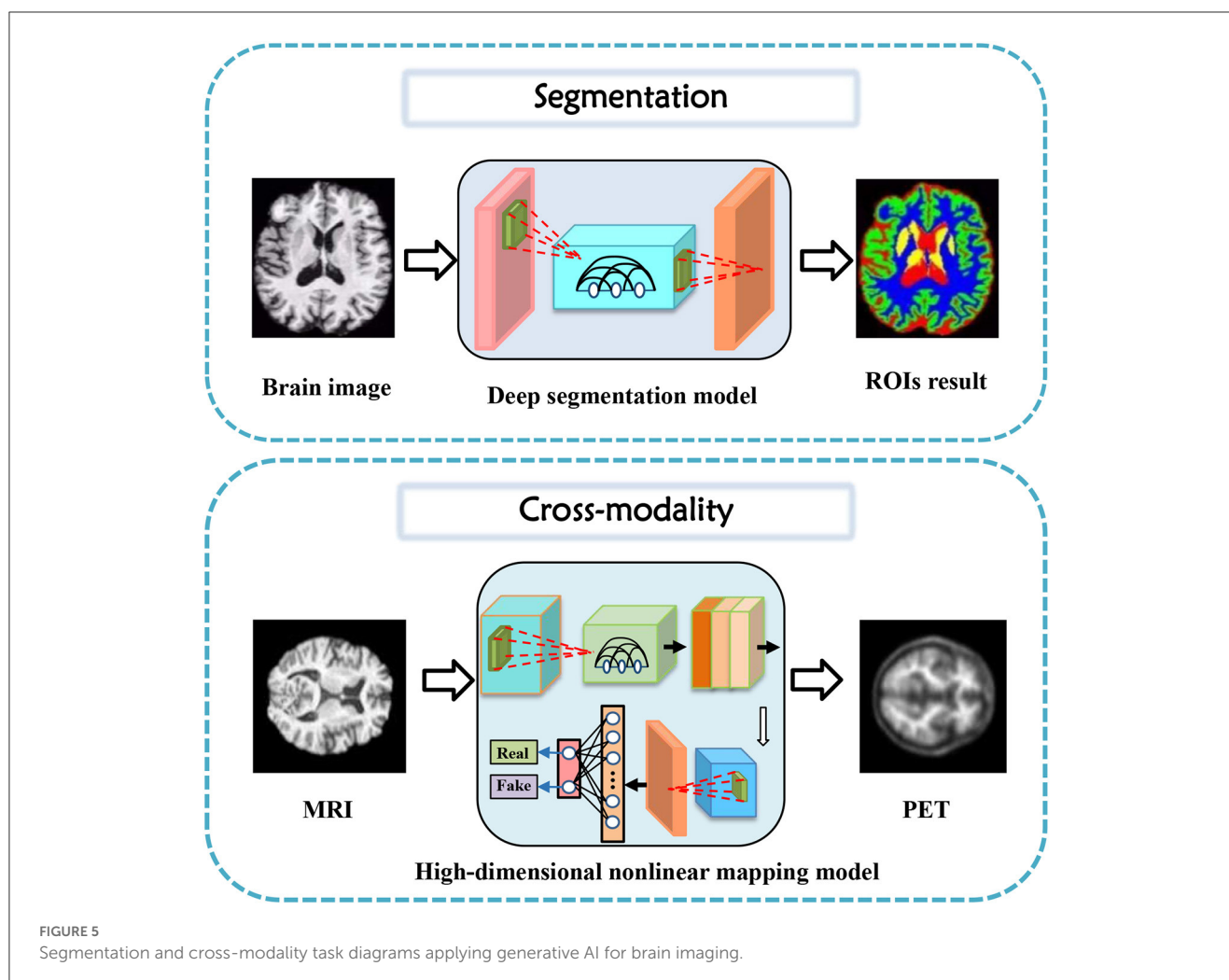
Alzheimer's disease (AD) and normal cognitive condition (NC). The authors use slice-selective learning to reduce computational costs and extract unbiased features. The researchers trained the model using an 18F-fluorodeoxyglucose ([18F] FDG) PET/CT dataset obtained from the Alzheimer's Disease Neuroimaging Initiative database. The approach seems feasible when there are insufficient datasets available for training individual deep neural networks (Wang et al., 2018; Yu et al., 2020) with single-source training datasets.

Furthermore, like Baur et al. (2020) propose a practical method for unsupervised brain MRI anomaly detection in clinical scenarios using a CycleGAN-based style-transfer framework. The proposed approach involves mapping real healthy data to a distribution with lower entropy and suppressing anomalies by filtering high-frequency components during training. The experiments demonstrate that the proposed method outperforms existing methods on various metrics, such as F1 score, PRC AUC, and ROC AUC, thus demonstrating its potential for practical applications in clinical settings.

Therefore, generative models have demonstrated their potential in various classification tasks for brain diseases. These models can extract features that are not directly visible, thereby aiding in the early diagnosis and accurate classification of diseases.

3.5. Segmentation

Generative models have gained significant attention in the field of medical image segmentation for their capability of reducing the dependence on manually labeled data. This paper reviewed the recent advances in generative models for segmentation tasks, focusing on brain tumor segmentation (Myronenko, 2018). In 2019, Yuan et al. (2020) presented a 3D unified generative adversarial network, achieving any-to-any modality translation and multimodal segmentation through a single network based on the anatomical structure. Ding et al. (2021) proposed ToStaGAN, a two-stage generative adversarial neural network, for brain tumor segmentation, which incorporates coarse prediction maps with fine-grained extraction modules and dense skip connections. In 2022, Wang S. et al. (2022) introduced Consistent Perception Generative Adversarial Network (CPGAN), an alternative to deep learning algorithms with expensive labeled masks, demonstrating superior segmentation performance over other methods with less labeled data on Anatomical Tracings of Lesions After Stroke. Wu et al. (2021) presented an unsupervised brain tumor segmentation method called Symmetric-Driven Generative Adversarial Network (SD-GAN) in 2021, which utilizes inherent anatomical variations by learning a non-linear mapping between



left and right brain images. SD-GAN outperforms state-of-the-art unsupervised methods, providing a promising solution to unsupervised segmentation tasks.

These studies demonstrate that generative models have become increasingly important for medical image segmentation owing to their ability to learn from unannotated data and promising performance in comparison to traditional supervised methods.

3.6. Cross-modality

Cross-modality image synthesis has become an active research area in medical imaging, where the goal is to generate images in a target modality from the input in another modality. Several generative models have been proposed for this task, including variations of generative adversarial networks (GANs) and encoder-decoder models. In recent years, significant progress has been made in using generative models for cross-modality image synthesis in the brain.

One of the early works in this area introduced gEa-GAN and dEa-GAN by Yu et al. (2019), which integrated edge information to bridge the gap between different imaging modalities. The

resulting synthesized images showed superior quality compared to several state-of-the-art methods, as demonstrated on various datasets. Another study (Hu et al., 2021) introduced Bidirectional GAN, which used a bidirectional mapping mechanism to embed diverse brain structural features into the high-dimensional latent space. The method achieved better quality in generating PET images than other models trained on the same dataset, while preserving diverse details of brain structures across different subjects. Other studies explored more challenging scenarios, such as Jiao et al. (2020) generating magnetic resonance (MR)-like images directly from clinical ultrasound (US) images of fetal brains and Sharma and Hamarneh (2019) synthesizing multiple modalities of neuroimaging data. The former study proposed an end-to-end trainable model that utilized shared latent features between US and MR data to generate realistic MR-like images, while the latter study introduced a multi-input, multi-output variant of GAN to synthesize sequences missing in brain MRI scans. The proposed models achieved promising results and demonstrate the feasibility of using generative models in clinical practice. Moreover, a bidirectional mapping mechanism is designed to embed the semantic information of PET images into the high-dimensional latent potential space for improving the visual quality of the

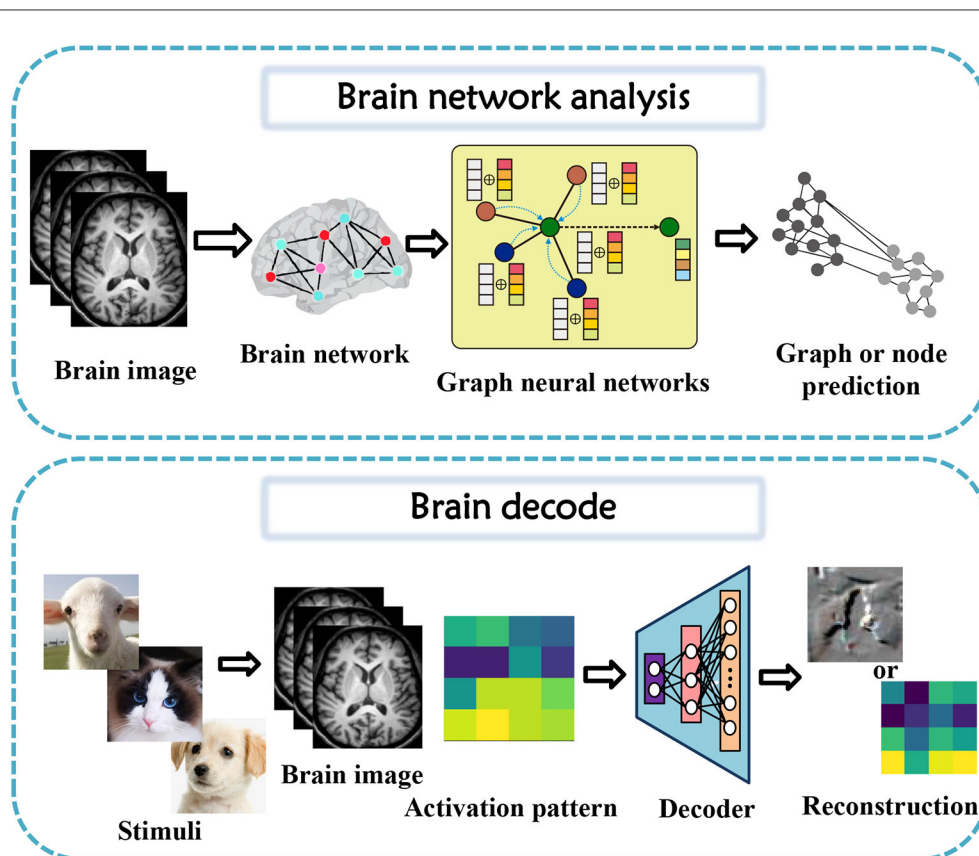


FIGURE 6

Brain network analysis and brain decode task diagrams applying generative AI for brain imaging.

cross-modal synthesized images (Hu et al., 2019, 2020a). The most attractive part is that the method can synthesize perceptually realistic PET images while preserving the different brain structures of different subjects.

Several other studies have also proposed novel generative models, including MouseGAN (Yu et al., 2021c) for segmenting mouse brain structures in MRI images and SC-GAN (Lan et al., 2021) for synthesizing multimodal 3D neuroimaging data. The former study achieved improved segmentation using modality-invariant information, while the latter used spectral normalization, feature matching, and self-attention modules to stabilize the training process and ensure optimization convergence. These studies have shown that generative models have the potential to improve existing neuroimaging analysis tasks and provide new tools for diagnosis and follow-up.

Finally, some studies have attempted to integrate generative models with disease diagnoses (Yang H. et al., 2021). Moreover, in neuroimaging data, One study (Pan et al., 2022) proposed a disease-image-specific deep learning framework that utilizes image-disease specificity to highlight different disease-relevant regions in the brain, with promising results on Alzheimer's Disease and mild cognitive impairment conversion prediction tasks. These studies highlight the potential of generative models to not only generate images of other modalities but also aid in downstream analysis and diagnoses using inter-modality information.

3.7. Brain network analysis

Brain network modeling is a critical research field in neuroscience that aims to understand the complex relationships among structural and functional connectivity patterns in the human brain. In recent years, deep learning has been increasingly used in brain network analysis as it shows promising results in predicting brain graphs, inferring effective connectivity, and diagnosing Alzheimer's disease using multimodal neuroimaging data. To this end, several deep learning frameworks have been proposed to generate reliable individual structural connectivity from functional connectivity, MultiGraphGAN, and MGCN-GAN proposed by Bessadok et al. (2020) and Zhang et al. (2020). These frameworks combine adversarial learning and topology preservation to generate high-quality brain graphs from limited data effectively. Some studies focused on inferring effective connectivity from functional MRI data, including EC-GAN and MGCN-GAN proposed by Liu et al. (2020), incorporate adversarial learning and graph convolutional networks to effectively infer brain structure-function relationships. Additionally, some studies aimed at diagnosing Alzheimer's disease using multimodal neuroimaging data, such as MRL-AHF and HGGAN proposed by Zuo et al. (2021) and Pan et al. (2021), which leverage adversarial learning and hypergraph representation to effectively integrate and represent multimodal

data. Overall, deep learning frameworks have been shown to hold great potential in brain network modeling, rapidly advancing our understanding of brain structure-function relationships and improving the prediction accuracy of brain disorders, such as brain drug addiction (Gong et al., 2023). Future research in this area will likely continue to explore and develop new deep learning-based approaches to further enhance modeling accuracy and generalization performance.

3.8. Brain decoding

Generative models have become a popular research focus in the field of brain decoding tasks, especially in the reconstruction of perceived images from fMRI signals. Baek et al. (2021) proposed a hierarchical deep neural network model of the ventral visual stream to explain the innate emergence of face-selectivity. VanRullen and Reddy (2019) applied a deep learning system to reconstruct face images from fMRI data, achieving accurate gender classification and decoding of visually similar inputs. Ren et al. (2021) proposed the Dual-Variational Autoencoder/Generative Adversarial Network framework, which outperforms state-of-the-art methods in terms of visual reconstruction accuracy. Chen et al. (2022) introduced the MinD-Vis framework, which uses a self-supervised representation of fMRI data and a latent diffusion model to reconstruct high-quality images with semantic details, outperforming state-of-the-art methods in semantic mapping and generation quality. Dado et al. (2022) presented a novel experimental paradigm, HYPER, for neural decoding of faces from brain recordings using generative adversarial networks, achieving the most accurate reconstructions of perception to date. These studies demonstrate the potential of generative models in brain decoding tasks, which can help advance our understanding of brain function and perception.

The application division of generative artificial intelligence methods in the field of brain image analysis is shown in Figure 7. The existing models mentioned above are divided according to tasks, and the names of the corresponding models are marked under the relevant tasks, which can be mainly divided into eight categories as shown in the figure. The most representative methods shown in the figure, the more research is optimized under this mainstream model. The four mainstream methods have different adaptation conditions for different tasks. In terms of coregistration, cross-modality, segmentation, classification, clustering, and super-scoring tasks, the optimization of the GAN model is significantly better than the other three mainstream models. The reason is that the GAN-based method can be well applied to brain image generation tasks. In brain network analysis and brain decoding tasks, the encoding and decoding structure of the VAE-based method will have more advantages. Flow-based models have relatively few applications and have a certain degree of application in super-resolution, brain decoding, and co-registration tasks. Diffusion models have high-quality generation effects and have been gradually used in various tasks of brain image analysis, and have achieved certain achievements.

4. Discussion

4.1. Challenge

The domain of brain image analysis and brain network computing confronted many obstacles that have impeded the development of the field. In order to progress future research more effectively, this review delves into some significant challenges and expounds on them in the following:

4.1.1. Small sample problem

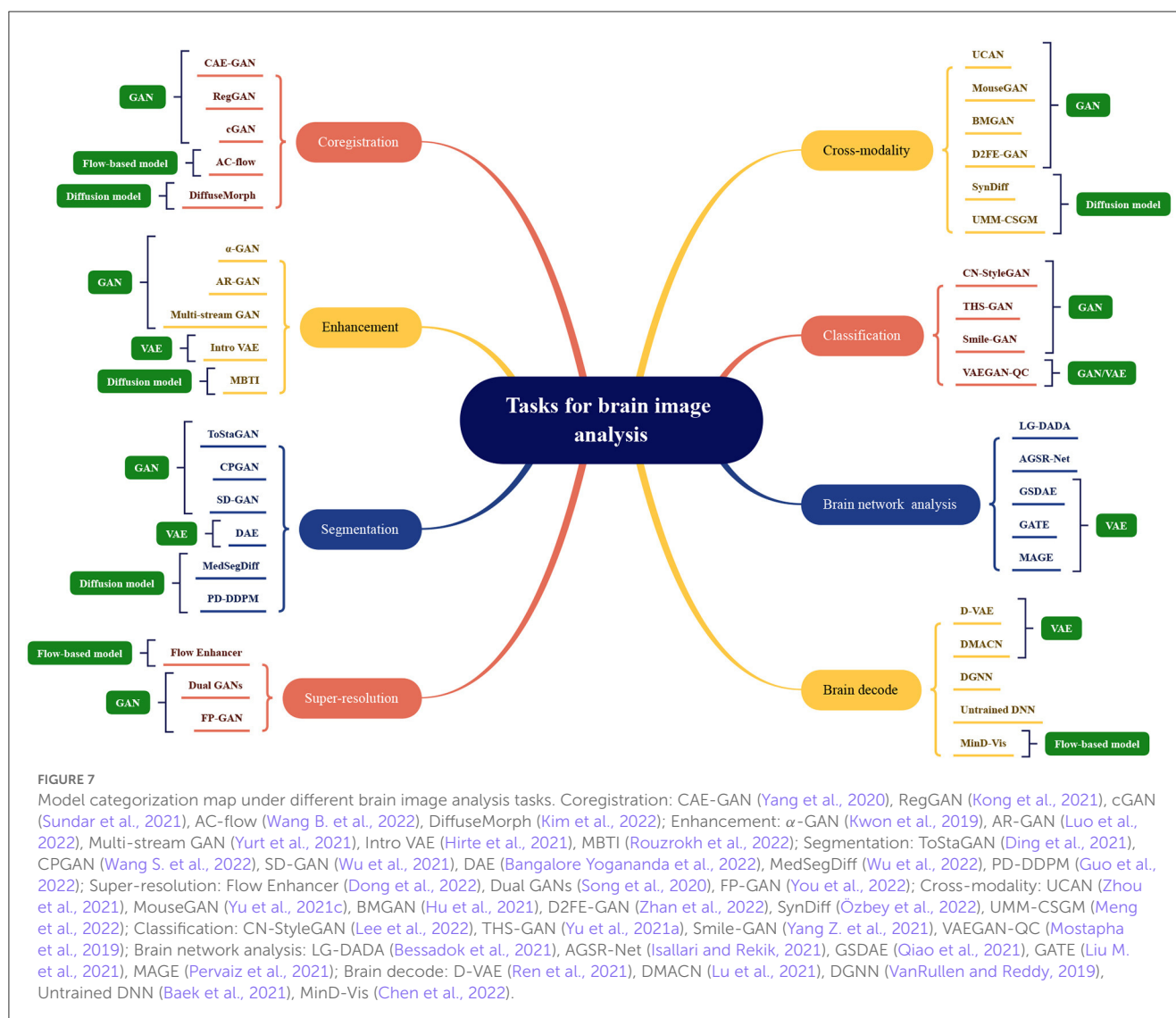
Medical image datasets are generally much smaller than datasets in other fields, due to the challenging task of acquiring and annotating medical images. For example, for lung nodule detection tasks, due to the small number of lung nodules, the number of positive samples in the dataset is very small, and the size, shape, and position of lung nodules also vary greatly, making it difficult for algorithms to accurately detect lung nodules. Therefore, the small sample problem has become one of the major obstacles for machine learning algorithms in the field of medical imaging. In addition to methods such as meta-learning that can effectively solve the small sample problem, improvement can be done by using prompts to modify pre-trained models, utilizing prior knowledge, and model ensembles. For example, in lung nodule detection tasks, pre-trained models can be used to extract features and prompts can be used to guide the model on how to detect lung nodules. In addition, prior knowledge can be used to constrain the output of the model, such as constraining the size and shape of the output lung nodules. Finally, model ensembles can be used to improve algorithm robustness and generalization capabilities.

4.1.2. High dimensional data problem

Medical scans or images are typically high-dimensional data types that contain large amounts of information. For example, in brain medical imaging, 3D MRI or fMRI is typically used to obtain brain structure and functional information. These data often contain millions of pixels or thousands of time points, so extracting meaningful, non-redundant, and non-overfitting features from such high-dimensional data is a new challenge for machine learning algorithms in the field of medical imaging. For the task of feature extraction from high-dimensional brain data, due to the complexity of brain structure and function, single modality feature extraction methods often have difficulty in extracting meaningful features. For example, in brain MRI images, tissue such as gray matter, white matter, and cerebrospinal fluid have different shapes and positions, so multi-modality feature analysis methods are required to extract meaningful features. In addition, as there are complex relationships between brain structure and function, multi-modality feature analysis methods are necessary to extract the relevant features between structure and function.

4.1.3. Realism of different modalities

Generative models have been widely applied in the field of medical data generation. However, the generated modality data may suffer from the problems of unreliability or modality



inconsistency. MRI and fMRI are common modalities in medical imaging. For MRI image generation, there may be problems such as insufficient reconstructed image resolution, lacking local details, and artifacts. In fMRI data generation, there may be signal suppression in local regions, interference from noise, and motion artifacts. In recent years, the generation of genetic data has also had a significant impact on the medical field. One major problem in generating genetic data is that the generated data may differ from real genetic data because real genetic data are produced by many genetic factors working together. Thus, when generating genetic data, multiple factors must be considered to improve the realism and consistency of generated data. While problems that could occur in extensively experimented image modalities may be easily identified and optimized, this realism problem can be hard to discern in complex or abstract modalities. Ultimately, such erroneously generated data may lead to incorrect diagnoses or treatments. To address this problem, researchers have proposed many methods, such as introducing special loss functions, such as cycle consistency in the model process in the image domain, and introducing strategies such as multi-modality

joint and multi-task learning to improve the generative quality and realism of modal data.

4.1.4. Standard validation issues

The lack of universally accepted validation standards for evaluating the performance of machine learning models in the field of medical imaging makes it difficult to compare results from different studies or ensure the best performance of the models. As the structure of the human brain is complex and diverse, specific to the generation task of brain diseases, more requirements are raised, and large differences exist between different regions, thus how to evaluate the performance of generative models to produce meaningful generative models. For example, for the generation task of MRI images, evaluation metrics can use traditional image quality evaluation indicators such as PSNR, SSIM, FID, etc., but these indicators cannot fully reflect the performance of the model in medical applications, such as whether the model generates anatomy structure consistent with real data, and whether it can better display lesion areas. Therefore, researchers have proposed

some specific evaluation indicators, such as similarity of structure with real images, neuron activation, diagnostic accuracy, etc., to more accurately evaluate the performance of generative models. However, the universality and comparability of these indicators still need more experimental verification and exploration in order to be better applied to different tasks and datasets.

4.1.5. Model interpretability

The issue of model interpretability is another challenge that must be addressed in medical image generation models. The model needs to have a certain level of interpretability so that physicians and researchers can understand the model's predictions and generate results. In the task of generating brain disease, model interpretability is particularly critical. Doctors need to be able to understand the relationship between the abnormal structures in the generated images and the underlying diseases, in order to make accurate diagnoses and treatment decisions. During the generation of images, the model may introduce factors such as image noise and artifacts, which can seriously affect medical diagnosis. To address these issues, interpretability techniques can provide valuable assistance. For example, visualization techniques can help doctors and researchers better understand the generated results and identify abnormal factors by generating comparative images and visualizing the internal feature maps of the model. In addition, model interpretability can also be achieved by adding interpretation layers or using interpretable models. However, there are some limitations to the use of data in the scenario of generating brain diseases. For example, due to privacy issues involving patients, medical imaging data is often highly sensitive and therefore difficult to obtain large-scale datasets directly. In addition, the different brain structures of different patients pose a challenge to the generated results of the model. Therefore, when generating brain diseases, it is necessary to consider the balance between data usage scenarios and model interpretability in order to obtain more accurate and interpretable results.

4.1.6. Limitation

There are still inescapable limitations to brain imaging computing using generative AI. Firstly, before the application of the synthetic brain image data set for training, if the differences between the synthetic data set and the real data set are not fully studied, the generated results will be biased. Secondly, most current generative brain image analysis methods may generate illogical "unnatural data" due to the lack of labels and causal features in the generating process. The robustness and reliability of the algorithm may be impacted. Thirdly, in the process of model training, it is possible to remember the distribution of original training samples. If the original training sample can be reversely inferred from the synthesized data, there will be "implicit privacy" leakage problem, and how to protect privacy more closely is still a question to be explored.

4.2. Future direction

New generative methods in brain network research are likely to find applications in both basic and clinical research. In the

coming years, generative learning and signal processing techniques will remain essential tools for furthering our understanding of the brain. This paper presents three perspectives on future approaches to brain network research:

4.2.1. Brain circuit identification

The application of generative models in brain circuit research is gaining more attention. Through the utilization of generative models, researchers are able to extract significant information about neural circuits from brain data. In the field of neuroscience, neural circuits are key elements for understanding brain function and are crucial for regulating various cognitive and affective behaviors. Deep learning can assist in revealing the intricate structure, profound functionality, and impact of the brain. Generative models can support neural circuit research by learning data features in brain circuits. In particular, generative models can be employed to generate, transform, and improve neuroimaging data, consequently creating novel and high-quality data to facilitate a more profound comprehension of neural circuits. Furthermore, generative models can serve as a data augmentation technique to diversify the limited neural circuit data samples, which can boost both training and disease diagnosis efficiency. Moreover, this technique can simulate and generate experiments related to the connection and variation of neural circuits among patients, ultimately resulting in more experimental evidence and predictive capacity for neural circuit research, therefore yielding more information for disease diagnosis and treatment.

4.2.2. Precise localization of brain regions

In the field of brain disorders, exploration of potential treatment options is a common practice. Generative models are capable of assisting in the comprehension of neural regulation and localization by producing intricate images that reflect the interconnectivity between different regions of the brain. Neural regulation denotes the procedure through which the brain controls behavior and emotions by moderating the excitatory and inhibitory activities of neurons. In neuroscience, generative models can simulate and forecast complex neural regulation processes. These models are instrumental in comprehending the mechanisms of neural regulation during brain development, deducing the reciprocal interaction between neurons and synapses, and predicting the patterns of connectivity between distinct neurons. Furthermore, synthetic neural imaging data is useful in providing researchers with a better appreciation of macroscopic neural regulation patterns. To pinpoint the location and function of particular brain areas, researchers can input significant image data into a generative model to learn the features of brain structures and more precisely localize brain regions. In functional connectivity analysis, generative models can generate hypothetical functional connectivity data and compare it with actual data to identify the links between various brain regions. The identification of possible targets for the treatment of brain-related disorders involves investigating the associations between the functions of neurons, synapses, and specific diseases. Generative models, as tools for data analysis and prediction, can effectively learn features and patterns automatically from vast amounts of neural data, and aid in target identification for the moderation or treatment of such diseases.

Thus, generative models can enable the comprehension of neural regulation and brain localization, facilitate the search for targets and solutions to treat brain disorders, and ultimately improve patients' quality of life.

4.2.3. Brain diseases and mechanisms

Neurodegenerative diseases, particularly Alzheimer's disease, are caused by damage to neurons and synapses in the brain, resulting in declines in cognitive and memory function and eventually leading to dementia. Studies have shown that generative models can forecast the speed of a patient's cognitive function decline, aiding physicians in early disease diagnosis. These models can also assimilate prior knowledge on diverse physiological, neural, cognitive, and behavioral processes, such as visual information processing, perception, language, and memory cognition. Generative models have helped in analyzing the onset mechanism of the disease and developing individualized treatment plans. Consequently, they are expected to become pivotal tools for exploring and comprehending brain mechanisms, with the potential to boost the precision of neurodegenerative disease diagnosis and treatment (Jing et al., 2022).

4.2.4. Multi-scale brain atlas

Population-based multi-scale brain research is a prominent focus in neuroscience (Betz et al., 2017). Its goal is to integrate information from multiple levels to understand the structure and function of the brain, establish connections between them, and gain insights into the working principles of the brain. Macro-level research investigates the overall structure and function of the brain, while micro-level research focuses on the cellular-level structure and function of neurons and synapses. Meso-scale research investigates small structures, such as cortical columns, connections, and neuronal clusters. Genomics studies the influence of genes on the brain's structure and function. Generative models integrate multiple data sources to reveal the complex mechanisms of the brain and explore the interactions between neurons and brain regions, providing a comprehensive view of the brain network. They can also predict gene expression data, diagnose and treat individual differences in diseases. Although the use of generative models in the study of the brain is still in its early stages, the accumulation of data and technological advances are expected to expand their usage.

These approaches will help build a comprehensive and accurate representation of the human brain and enable the discovery of new insights across neurological and psychiatric disorders.

5. Conclusion

This article provides a review of generative artificial intelligence for brain image computing and brain network computing. Generative AI can be divided into four main methods: variational autoencoder (VAE), generative adversarial network (GAN), flow-based model, and diffusion model. These models offer a promising

solution for analyzing and interpreting large-scale brain imaging data. Generative AI has enabled researchers to gain a better understanding of the brain's physical basis and how it adapts to various cognitive activities in the field of brain imaging. In the context of brain network computing, generative AI can be used to reconstruct the topological connectivity of brain networks. However, there are limitations associated with using generative AI for analyzing brain imaging data. For instance, medical imaging data is often highly sensitive due to privacy issues involving patients, making it difficult to obtain large-scale datasets directly. Additionally, different brain structures of different patients pose a challenge to the generated results of the model. Therefore, when using generative AI techniques to generate brain diseases or analyze large-scale medical imaging datasets, it is necessary to balance data usage scenarios and model interpretability in order to obtain more accurate and interpretable results. In conclusion, generative AI has broad application prospects in brain imaging and brain network, which can help to better understand the internal function and structure of the brain, promote the diagnosis and treatment of brain diseases, and provide new opportunities and methods for neuroscience research.

Author contributions

SW and YH proposed the idea and co-managed the project. YH, SW, and CG co-designed the framework. CG, CJ, XC, CP, GH, SW, and YH co-wrote the manuscript. MN and AS supported the project and offered resources to accomplish this research. All authors read, contributed to revision, and approved the manuscript.

Funding

This work was supported by the National Natural Science Foundations of China under Grant 62172403, the Distinguished Young Scholars Fund of Guangdong under Grant 2021B1515020019, the Excellent Young Scholars of Shenzhen under Grant RCYX20200714114641211, and Shenzhen Key Basic Research Projects under Grant JCYJ20200109115641762.

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

- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352
- Arjovsky, M., Chintala, S., and Bottou, L. (2017). “Wasserstein generative adversarial networks,” in *International Conference on Machine Learning* (Sydney: PMLR), 214–223.
- Baek, S., Song, M., Jang, J., Kim, G., and Paik, S.-B. (2021). Face detection in untrained deep neural networks. *Nat. Commun.* 12, 1–15. doi: 10.1038/s41467-021-27606-9
- Bangalore Yogananda, C. G., Das, Y., Wagner, B. C., Nalawade, S. S., Reddy, D., Holcomb, J., et al. (2022). “Disparity autoencoders for multi-class brain tumor segmentation,” in *International MICCAI Brainlesion Workshop* (Singapore: Springer), 116–124.
- Bassett, D. S., and Sporns, O. (2017). Network neuroscience. *Nat. Neurosci.* 20, 353–364. doi: 10.1038/nn.4502
- Baur, C., Graf, R., Wiestler, B., Albarqouni, S., and Navab, N. (2020). “Steganomaly: inhibiting cyclegan steganography for unsupervised anomaly detection in brain MRI,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Lima: Springer), 718–727. doi: 10.1007/978-3-030-59713-9_69
- Bessadok, A., Mahjoub, M. A., and Reik, I. (2020). “Topology-aware generative adversarial network for joint prediction of multiple brain graphs from a single brain graph,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Lima: Springer), 551–561.
- Bessadok, A., Mahjoub, M. A., and Reik, I. (2021). Brain graph synthesis by dual adversarial domain alignment and target graph prediction from a source graph. *Med. Image Anal.* 68, 101902. doi: 10.1016/j.media.2020.101902
- Betz, R. F., and Bassett, D. S. (2017). Multi-scale brain networks. *Neuroimage* 160, 73–83. doi: 10.1016/j.neuroimage.2016.11.006
- Calhoun, V. D., Miller, R., Pearlson, G., and Adalı, T. (2014). The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* 84, 262–274. doi: 10.1016/j.neuron.2014.10.015
- Chen, Z., Qing, J., Xiang, T., Yue, W. L., and Zhou, J. H. (2022). Seeing beyond the brain: conditional diffusion model with sparse masked modeling for vision decoding. *arXiv*.
- Cheng, J., Gao, M., Liu, J., Yue, H., Kuang, H., Liu, J., et al. (2021). Multimodal disentangled variational autoencoder with game theoretic interpretability for glioma grading. *IEEE J. Biomed. Health Informat.* 26, 673–684. doi: 10.1109/JBHI.2021.3095476
- Chung, H., and Ye, J.-C. (2022). Score-based diffusion models for accelerated MRI. *Med. Image Anal.* 80, 102479. doi: 10.1016/j.media.2022.102479
- Cui, Z., Zhong, S., Xu, P., He, Y., and Gong, G. (2013). Panda: a pipeline toolbox for analyzing brain diffusion images. *Front. Hum. Neurosci.* 7, 42. doi: 10.3389/fnhum.2013.00042
- Dado, T., Güçlütürk, Y., Ambrogioni, L., Ras, G., Bosch, S., van Gerven, M., et al. (2022). Hyperrealistic neural decoding for reconstructing faces from fMRI activations via the GAN latent space. *Sci. Rep.* 12, 1–9. doi: 10.1038/s41598-021-03938-w
- Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., and Çukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans. Med. Imaging* 38, 2375–2388. doi: 10.1109/TMI.2019.2901750
- Ding, Y., Zhang, C., Cao, M., Wang, Y., Chen, D., Zhang, N., et al. (2021). Tostagan: an end-to-end two-stage generative adversarial network for brain tumor segmentation. *Neurocomputing* 462, 141–153. doi: 10.1016/j.neucom.2021.07.066
- Dong, S., Hangel, G., Chen, E. Z., Sun, S., Bogner, W., Widhalm, G., et al. (2022). “Flow-based visual quality enhancer for super-resolution magnetic resonance spectroscopic imaging,” in *MICCAI Workshop on Deep Generative Models* (Springer), 3–13.
- Gao, X., Shi, F., Shen, D., and Liu, M. (2021). Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in Alzheimers disease. *IEEE J. Biomed. Health Informat.* 26, 36–43. doi: 10.1109/JBHI.2021.3097721
- Gong, C., Chen, X., Mughal, B., and Wang, S. (2023). Addictive brain-network identification by spatial attention recurrent network with feature selection. *Brain Informat.* 10, 1–11. doi: 10.1186/s40708-022-00182-4
- Gong, C., Xue, B., Jing, C., He, C.-H., Wu, G.-C., Lei, B., et al. (2022). Time-sequential graph adversarial learning for brain modularity community detection. *Math. Biosci. Eng.* 19, 13276–13293. doi: 10.3934/mbe.2022621
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, eds Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Montreal, QC: Curran Associates, Inc.), 2672–2680.
- Gui, X., Chuansheng, C., Zhong-Lin, L., and Qi, D. (2010). Brain imaging techniques and their applications in decision-making research. *Xin Li Xue Bao* 42, 120. doi: 10.3724/SP.J.1041.2010.00120
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. *Adv. Neural Inf. Process. Syst.* 30, 5769–5779. doi: 10.5555/3295222.3295327
- Guo, X., Yang, Y., Ye, C., Lu, S., Xiang, Y., and Ma, T. (2022). Accelerating diffusion models via pre-segmentation diffusion sampling for medical image segmentation. *arXiv*.
- Hirte, A. U., Platscher, M., Joyce, T., Heit, J. J., Tranvinh, E., and Federau, C. (2021). Realistic generation of diffusion-weighted magnetic resonance brain images with deep generative models. *Magn. Reson. Imaging* 81, 60–66. doi: 10.1016/j.mri.2021.06.001
- Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* 33, 6840–6851. doi: 10.5555/3495724.3496298
- Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., and Shen, Y. (2021). Bidirectional mapping generative adversarial networks for brain mr to pet synthesis. *IEEE Trans. Med. Imaging* 41, 145–157. doi: 10.1109/TMI.2021.3107013
- Hu, S., Shen, Y., Wang, S., and Lei, B. (2020a). “Brain MR to pet synthesis via bidirectional generative adversarial network,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer), 698–707.
- Hu, S., Yu, W., Chen, Z., and Wang, S. (2020b). “Medical image reconstruction using generative adversarial network for Alzheimer’s disease assessment with class-imbalance problem,” in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)* (Chengdu: IEEE), 1323–1327.
- Hu, S., Yuan, J., and Wang, S. (2019). “Cross-modality synthesis from MRI to pet using adversarial u-net with different normalization,” in *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)* (IEEE), 1–5.
- Isallari, M., and Reik, I. (2021). Brain graph super-resolution using adversarial graph neural network with application to functional brain connectivity. *Med. Image Anal.* 71, 102084. doi: 10.1016/j.media.2021.102084
- Jiao, J., Namburete, A. I. L., Papageorgiou, A. T., and Noble, J. A. (2020). Self-supervised ultrasound to MRI fetal brain image synthesis. *IEEE Trans. Med. Imaging* 39, 4413–4424. doi: 10.1109/TMI.2020.3018560
- Jing, C., Gong, C., Chen, Z., Lei, B., and Wang, S. (2022). TA-GAN: transformer-driven addition-perception generative adversarial network. *Neural Comp. Appl.* 35, 1–13. doi: 10.1007/s00521-022-08187-0
- Joyce, J. M. (2011). “Kullback-leibler divergence,” in *International Encyclopedia of Statistical Science* (Springer), 720–722.
- Kim, B., Han, I., and Ye, J. C. (2022). “Diffusemorph: Unsupervised deformable image registration using diffusion model,” in *European Conference on Computer Vision* (Glasgow: Springer), 347–364.
- Kim, H. W., Lee, H. E., Lee, S., Oh, K. T., Yun, M., and Yoo, S. K. (2020). Slice-selective learning for Alzheimer’s disease classification using a generative adversarial network: a feasibility study of external validation. *Eur. J. Nucl. Med. Mol. Imag.* 47, 2197–2206. doi: 10.1007/s00259-019-04676-y
- Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, J. (2017). “Learning to discover cross-domain relations with generative adversarial networks,” in *International Conference on Machine Learning* (Sydney, NSW: PMLR), 1857–1865.
- Kingma, D. P., and Welling, M. (2013). Auto-encoding variational bayes. *arXiv*.
- Kong, L., Lian, C., Huang, D., Hu, Y., and Zhou, Q. (2021). Breaking the dilemma of medical image-to-image translation. *Adv. Neural Inf. Process. Syst.* 34, 1964–1978.
- Kwon, G., Han, C., and Kim, D.-S. (2019). “Generation of 3D brain MRI using auto-encoding generative adversarial networks,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 118–126.
- Lan, H., Initiative, A. D. N., Toga, A. W., and Sepreband, F. (2021). Three-dimensional self-attention conditional GAN with spectral normalization for multimodal neuroimaging synthesis. *Magn. Resonan. Med.* 86, 1718–1733. doi: 10.1002/mrm.28819
- Lee, S., Jeong, B., Kim, M., Jang, R., Paik, W., Kang, J., et al. (2022). Emergency triage of brain computed tomography via anomaly detection with a deep generative model. *Nat. Commun.* 13, 1–11. doi: 10.1038/s41467-022-31808-0
- Lei, B., Liang, E., Yang, M., Yang, P., Zhou, F., Tan, E.-L., et al. (2022a). Predicting clinical scores for Alzheimer’s disease based on joint and deep learning. *Expert Syst. Appl.* 187, 115966. doi: 10.1016/j.eswa.2021.115966
- Lei, B., Xia, Z., Jiang, F., Jiang, X., Ge, Z., Xu, Y., et al. (2020a). Skin lesion segmentation via generative adversarial networks with dual discriminators. *Med. Image Anal.* 64, 101716. doi: 10.1016/j.media.2020.101716
- Lei, B., Yang, M., Yang, P., Zhou, F., Hou, W., Zou, W., et al. (2020b). Deep and joint learning of longitudinal data for Alzheimer’s disease prediction. *Pattern Recognit.* 102, 107247. doi: 10.1016/j.patcog.2020.107247

- Lei, B., Zhang, Y., Liu, D., Xu, Y., Yue, G., Cao, J., et al. (2022b). Longitudinal study of early mild cognitive impairment via similarity-constrained group learning and self-attention based SBI-LSTM. *Knowledge-Based Systems* 254:109466. doi: 10.1016/j.knsys.2022.109466
- Liu, J., Ji, J., Xun, G., Yao, L., Huai, M., and Zhang, A. (2020). EC-GAN: inferring brain effective connectivity via generative adversarial networks. *Proc. AAAI Conf. Artif. Intell.* 34, 4852–4859. doi: 10.1609/aaai.v34i04.5921
- Liu, M., Zhang, Z., and Dunson, D. B. (2021). Graph auto-encoding brain networks with applications to analyzing large-scale brain imaging datasets. *Neuroimage* 245, 118750–118750. doi: 10.1016/j.neuroimage.2021.118750
- Liu, M.-Y., Breuel, T., and Kautz, J. (2017). Unsupervised image-to-image translation networks. *Adv. Neural Inf. Process. Syst.* 30, 700–708. doi: 10.1007/978-3-319-70139-4
- Liu, Y., Liu, J., and Wang, S. (2021). “Effective distributed learning with random features: Improved bounds and algorithms,” in *International Conference on Learning Representations* (Vienna).
- Lu, H., Liu, S., Wei, H., Chen, C., and Geng, X. (2021). Deep multi-kernel auto-encoder network for clustering brain functional connectivity data. *Neural Netw.* 135, 148–157. doi: 10.1016/j.neunet.2020.12.005
- Lundervold, A. S., and Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschr. Med. Phys.* 29, 102–127. doi: 10.1016/j.zemedi.2018.11.002
- Luo, Y., Nie, D., Zhan, B., Li, Z., Wu, X., Zhou, J., et al. (2021). Edge-preserving MRI image synthesis via adversarial network with iterative multi-scale fusion. *Neurocomputing* 452, 63–77. doi: 10.1016/j.neucom.2021.04.060
- Luo, Y., Zhou, L., Zhan, B., Fei, Y., Zhou, J., Wang, Y., et al. (2022). Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis. *Med. Image Anal.* 77, 102335. doi: 10.1016/j.media.2021.102335
- Meng, X., Gu, Y., Pan, Y., zhuan Wang, N., Xue, P., Lu, M., et al. (2022). A novel unified conditional score-based generative framework for multi-modal medical image completion. *arXiv*.
- Mo, L., and Wang, S.-Q. (2009). A variational approach to nonlinear two-point boundary value problems. *Nonlinear Anal. Theory Methods Appl.* 71, 834–838. doi: 10.1016/j.na.2008.12.006
- Mostapha, M., Prieto, J. C., Murphy, V., Girault, J. B., Foster, M., Rumble, A., et al. (2019). “Semi-supervised VAE-GAN for out-of-sample detection applied to MRI quality control,” in *MICCAI* (Shenzhen).
- Myronenko, A. (2018). “3D MRI brain tumor segmentation using autoencoder regularization,” in *International MICCAI Brainlesion Workshop* (Granada: Springer), 311–320.
- Özbey, M., Dalmaz, O., Dar, S. U., Bedel, H. A., Şaban Öztürk, Güngör, A., and Çukur, T. (2022). Unsupervised medical image translation with adversarial diffusion models. *arXiv*.
- Pan, J., Lei, B., Shen, Y., Liu, Y., Feng, Z., and Wang, S. (2021). “Characterization multimodal connectivity of brain network by hypergraph GAN for Alzheimer’s disease analysis,” in *PRCV2021*, 467–478.
- Pan, Y., Liu, M., Lian, C., Xia, Y., and Shen, D. (2019). “Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multimodal neuroimages,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 137–145.
- Pan, Y., Liu, M., Xia, Y., and Shen, D. (2022). Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 6839–6853. doi: 10.1109/TPAMI.2021.3091214
- Pervaiz, U., Vidaurre, D., Gohil, C., Smith, S. M., and Woolrich, M. W. (2021). Multi-dynamic modelling reveals strongly time-varying resting fMRI correlations. *Med. Image Anal.* 77, 102366. doi: 10.1101/2021.06.23.449584
- Qiao, C., Hu, X.-Y., Xiao, L., Calhoun, V. D., and Wang, Y.-P. (2021). A deep autoencoder with sparse and graph laplacian regularization for characterizing dynamic functional connectivity during brain development. *Neurocomputing* 456, 97–108. doi: 10.1016/j.neucom.2021.05.003
- Ren, Z., Li, J., Xue, X., Li, X., Yang, F., Jiao, Z., et al. (2021). Reconstructing seen image from brain activity by visually-guided cognitive representation and adversarial learning. *Neuroimage* 228, 117602. doi: 10.1016/j.neuroimage.2020.117602
- Rezende, D. J., and Mohamed, S. (2015). “Variational inference with normalizing flows,” in *International Conference on Machine Learning* (Lille: PMLR), 1530–1538.
- Rouzokh, P., Khosravi, P., Faghani, S., Moassefi, M., Vahdati, S., and Erickson, B. J. (2022). Multitask brain tumor inpainting with diffusion models: a methodological report. *arXiv*.
- Sharma, A., and Hamarneh, G. (2019). Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. *IEEE Trans. Med. Imaging* 39, 1170–1183. doi: 10.1109/TMI.2019.2945521
- Shin, H.-C., Ihsani, A., Xu, Z., Mandava, S., Sreenivas, S. T., Forster, C., et al. (2020). “Gandalf: Generative adversarial networks with discriminator-adaptive loss fine-tuning for Alzheimer’s disease diagnosis from MRI,” in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020* (Cham: Springer International Publishing), 688–697.
- Song, T.-A., Chowdhury, S. R., Yang, F., and Dutta, J. (2020). Pet image super-resolution using generative adversarial networks. *Neural Netw.* 125, 83–91. doi: 10.1016/j.neunet.2020.01.029
- Sun, L., Chen, J., Xu, Y., Gong, M., Yu, K., and Batmanghelich, K. (2022). Hierarchical amortized GAN for 3D high resolution medical image synthesis. *IEEE J. Biomed. Health Informat.* 26, 3966–3975. doi: 10.1109/JBHI.2022.3172976
- Sundar, L. K. S., Iommi, D., Muzik, O., Chalampalakakis, Z., Klebermass, E.-M., Hienert, M., et al. (2021). Conditional generative adversarial networks aided motion correction of dynamic 18F-FDG pet brain studies. *J. Nucl. Med.* 62, 871–879. doi: 10.2967/jnumed.120.248856
- Tezcan, K. C., Baumgartner, C. F., Luechinger, R., Pruessmann, K. P., and Konukoglu, E. (2017). MR image reconstruction using deep density priors. *IEEE Trans. Med. Imag.* 38, 1633–1642. doi: 10.1109/TMI.2018.2887072
- Upadhyay, U., Sudarshan, V. P., and Awate, S. P. (2021). “Uncertainty-aware GAN with adaptive loss for robust MRI image enhancement,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Montreal, QC), 3255–3264.
- VanRullen, R., and Reddy, L. (2019). Reconstructing faces from fMRI patterns using deep generative neural networks. *Commun. Biol.* 2, 1–10. doi: 10.1038/s42003-019-0438-y
- Wang, B., Lu, L., and Liu, H. (2022). “Invertible ac-flow: direct attenuation correction of pet images without ct or mr images,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)* (Kolkata: IEEE), 1–4.
- Wang, J., Wang, X., Xia, M., Liao, X., Evans, A., and He, Y. (2015). GRETN: a graph theoretical network analysis toolbox for imaging connectomics. *Front. Hum. Neurosci.* 9, 386. doi: 10.3389/fnhum.2015.00386
- Wang, S., Chen, Z., You, S., Wang, B., Shen, Y., and Lei, B. (2022). Brain stroke lesion segmentation using consistent perception generative adversarial network. *Neural Comp. Appl.* 34, 8657–8669. doi: 10.1007/s00521-021-06816-8
- Wang, S., Shen, Y., Chen, W., Xiao, T., and Hu, J. (2017). “Automatic recognition of mild cognitive impairment from MRI images using expedited convolutional neural networks,” in *Artificial Neural Networks and Machine Learning-ICANN 2017: 26th International Conference on Artificial Neural Networks, Proceedings, Part I 26* (Alghero: Springer), 373–380.
- Wang, S., Shen, Y., Zeng, D., and Hu, Y. (2018). “Bone age assessment using convolutional neural networks,” in *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)* (Chengdu: IEEE), 175–178.
- Wang, S., Wang, H., Cheung, A. C., Shen, Y., and Gan, M. (2020a). “Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer’s disease,” in *Deep Learning Applications* (Singapore: Springer), 53–73.
- Wang, S., Wang, X., Hu, Y., Shen, Y., Yang, Z., Gan, M., et al. (2020b). Diabetic retinopathy diagnosis using multichannel generative adversarial network with semisupervision. *IEEE Transact. Automat. Sci. Eng.* 18, 574–585. doi: 10.1109/TASE.2020.2981637
- Wang, S.-Q., Li, X., Cui, J.-L., Li, H.-X., Luk, K. D., and Hu, Y. (2015). Prediction of myelopathic level in cervical spondylotic myelopathy using diffusion tensor imaging. *J. Magn. Reson. Imag.* 41, 1682–1688. doi: 10.1002/jmri.24709
- Wicaksono, K. P., Fujimoto, K., Fushimi, Y., Sakata, A., Okuchi, S., Hinoda, T., et al. (2022). Super-resolution application of generative adversarial network on brain time-of-flight mr angiography: image quality and diagnostic utility evaluation. *Eur. Radiol.* 33, 1–11. doi: 10.1007/s00330-022-09103-9
- Wolleb, J., Bieder, F., Sandkühler, R., and Cattin, P. C. (2022). “Diffusion models for medical anomaly detection,” in *MICCAI* (Singapore).
- Wolterink, J. M., Dinkla, A. M., Savenije, M., Seevinck, P. R., van den Berg, C., and Išgum, I. (2017). “MR-to-CT synthesis using cycle-consistent generative adversarial networks,” in *Proc. Neural Inf. Process. Syst.* (Long Beach, CA: NIPS).
- Wu, J., Fang, H., Zhang, Y., Yang, Y., and Xu, Y. (2022). Medsegdiff: Medical image segmentation with diffusion probabilistic model. *arXiv*.
- Wu, X., Bi, L., Fulham, M., Feng, D. D., Zhou, L., and Kim, J. (2021). Unsupervised brain tumor segmentation using a symmetric-driven adversarial network. *Neurocomputing* 455, 242–254. doi: 10.1016/j.neucom.2021.05.073
- Yang, H., Qian, P., and Fan, C. (2020). An indirect multimodal image registration and completion method guided by image synthesis. *Comput. Math. Methods Med.* 2020, 2684851. doi: 10.1155/2020/2684851
- Yang, H., Sun, J., Yang, L., and Xu, Z. (2021). “A unified hyper-GAN model for unpaired multi-contrast mr image translation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Strasbourg: Springer), 127–137.
- Yang, Z., Nasrallah, I. M., Shou, H., Wen, J., Doshi, J., Habes, M., et al. (2021). A deep learning framework identifies dimensional representations of Alzheimer’s disease from brain structure. *Nat. Commun.* 12, 1–15. doi: 10.1038/s41467-021-26703-z
- Yi, Z., Zhang, H., Tan, P., and Gong, M. (2017). “Dualgan: unsupervised dual learning for image-to-image translation,” in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 2849–2857.

- You, S., Lei, B., Wang, S., Chui, C. K., Cheung, A. C., Liu, Y., et al. (2022). Fine perceptive gans for brain MR image super-resolution in wavelet domain. *IEEE Transac. Neural Netw. Learn. Syst.* 1–13. doi: 10.1109/TNNLS.2022.3153088
- Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., and Bourgeat, P. (2019). EA-GANS: edge-aware generative adversarial networks for cross-modality MR image synthesis. *IEEE Trans. Med. Imaging* 38, 1750–1762. doi: 10.1109/TMI.2019.2895894
- Yu, S., Wang, S., Xiao, X., Cao, J., Yue, G., Liu, D., et al. (2020). “Multi-scale enhanced graph convolutional network for early mild cognitive impairment detection,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Lima: Springer), 228–237.
- Yu, W., Lei, B., Ng, M. K., Cheung, A. C., Shen, Y., and Wang, S. (2021a). “Tensorizing GAN with high-order pooling for Alzheimer’s disease assessment,” in *IEEE Transactions on Neural Networks and Learning Systems*.
- Yu, W., Lei, B., Shen, Y., Wang, S., Liu, Y., Feng, Z., et al. (2021b). “Morphological feature visualization of Alzheimer’s disease via multidirectional perception GAN,” in *IEEE Transactions on Neural Networks and Learning Systems*.
- Yu, Z., Zhai, Y., Han, X., Peng, T., and Zhang, X.-Y. (2021c). “Mousegan: GAN-based multiple MRI modalities synthesis and segmentation for mouse brain structures,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Strasbourg: Springer), 442–450.
- Yuan, W., Wei, J., Wang, J., Ma, Q., and Tasdizen, T. (2020). Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images. *Med. Image Anal.* 64, 101731. doi: 10.1016/j.media.2020.101731
- Yurt, M., Dar, S. U., Erdem, A., Erdem, E., Oguz, K. K., and Çukur, T. (2021). MUSTGAN: multi-stream generative adversarial networks for mr image synthesis. *Med. Image Anal.* 70, 101944. doi: 10.1016/j.media.2020.101944
- Zhan, B., Li, D., Wu, X., Zhou, J., and Wang, Y. (2021). Multi-modal MRI image synthesis via GAN with multi-scale gate merge. *IEEE J. Biomed. Health Informat.* 26, 17–26. doi: 10.1109/JBHI.2021.3088866
- Zhan, B., Zhou, L., Li, Z., Wu, X., Pu, Y., Zhou, J., et al. (2022). D2FE-GAN: decoupled dual feature extraction based GAN for MRI image synthesis. *Knowl. Based Syst.* 252, 109362. doi: 10.1016/j.knsys.2022.109362
- Zhang, L., Wang, L., and Zhu, D. (2020). “Recovering brain structural connectivity from functional connectivity via multi-gcn based generative adversarial network,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Lima: Springer), 53–61.
- Zhao, Q., Adeli, E., Honnorat, N., Leng, T., and Pohl, K. M. (2019). “Variational autoencoder for regression: Application to brain aging analysis,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 823–831.
- Zhou, B., Wang, R., Chen, M.-K., Mecca, A. P., O’Dell, R. S., Dyck, C. H. V., et al. (2021). “Synthesizing multi-tracer pet images for Alzheimer’s disease patients using a 3D unified anatomy-aware cyclic adversarial network,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Strasbourg: Springer), 34–43.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 2223–2232.
- Zuo, Q., Lei, B., Shen, Y., Liu, Y., Feng, Z., and Wang, S. (2021). “Multimodal representations learning and adversarial hypergraph fusion for early Alzheimer’s disease prediction,” in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)* (Beijing: Springer), 479–490.



OPEN ACCESS

EDITED BY

Shuqiang Wang,
Chinese Academy of Sciences (CAS), China

REVIEWED BY

Feng Liu,
Stevens Institute of Technology, United States
Bor-Shing Lin,
National Taipei University, Taiwan

*CORRESPONDENCE

Ning Zhong
✉ zhong@maebashi-it.ac.jp
Jianzhao Yan
✉ yanjianzhao@bjut.edu.cn

[†]These authors have contributed equally to this work

RECEIVED 08 April 2023

ACCEPTED 09 May 2023

PUBLISHED 23 June 2023

CITATION

Cao Y, Kuai H, Liang P, Pan J-S, Yan J and Zhong N (2023) BNLoop-GAN: a multi-loop generative adversarial model on brain network learning to classify Alzheimer's disease. *Front. Neurosci.* 17:1202382. doi: 10.3389/fnins.2023.1202382

COPYRIGHT

© 2023 Cao, Kuai, Liang, Pan, Yan and Zhong. 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.

BNLoop-GAN: a multi-loop generative adversarial model on brain network learning to classify Alzheimer's disease

Yu Cao^{1,2†}, Hongzhi Kuai^{3†}, Peipeng Liang⁴, Jeng-Shyang Pan⁵, Jianzhao Yan^{1,2*} and Ning Zhong^{2,3,4*}

¹Faculty of Information Technology, Beijing University of Technology, Beijing, China, ²Beijing International Collaboration Base on Brain Informatics and Wisdom Services, Beijing, China, ³Faculty of Engineering, Maebashi Institute of Technology, Maebashi, Gunma, Japan, ⁴School of Psychology and Beijing Key Laboratory of Learning and Cognition, Capital Normal University, Beijing, China, ⁵College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China

Recent advancements in AI, big data analytics, and magnetic resonance imaging (MRI) have revolutionized the study of brain diseases such as Alzheimer's Disease (AD). However, most AI models used for neuroimaging classification tasks have limitations in their learning strategies, that is batch training without the incremental learning capability. To address such limitations, the systematic Brain Informatics methodology is reconsidered to realize evidence combination and fusion computing with multi-modal neuroimaging data through continuous learning. Specifically, we introduce the BNLoop-GAN (Loop-based Generative Adversarial Network for Brain Network) model, utilizing multiple techniques such as conditional generation, patch-based discrimination, and Wasserstein gradient penalty to learn the implicit distribution of brain networks. Moreover, a multiple-loop-learning algorithm is developed to combine evidence with better sample contribution ranking during training processes. The effectiveness of our approach is demonstrated through a case study on the classification of individuals with AD and healthy control groups using various experimental design strategies and multi-modal brain networks. The BNLoop-GAN model with multi-modal brain networks and multiple-loop-learning can improve classification performance.

KEYWORDS

BNLoop-GAN model, multiple-loop-learning, evidence combination-fusion computing, magnetic resonance imaging, brain network analysis, Alzheimer's disease

1. Introduction

The rapid advancement of AI and big data technologies have revolutionized the field of brain investigation, providing new insights into its workings and potential applications. However, medical research on the brain presents more significant challenges as it involves navigating the complex interplay of biological, psychological, and environmental factors. In response to this, the Brain Informatics (Zhong et al., 2011) methodology has been proposed to study the mechanisms underlying the human information processing system with big data (Zhong et al., 2005). As the core part of Brain Informatics, a series of "evidence combination-fusion computing (ECFC)" methods (Kuai et al., 2022) are developed to promote fundamental and translational studies of the brain, encouraging to handle multi-source brain big data continuously during

learning and validating phases of models and systems. The continuous learning enables the more effective utilization of existing information and experiences learned by previous data, which are different from the current most machine learning algorithms.

Alzheimer's Disease (AD), as a neurodegenerative disease that occurs frequently in the elderly, has become a severe threat to the health, with clinical manifestations of cognitive decline, accompanied by other physiological or mental disorders (Citron, 2010; Ferrari and Sorbi, 2021). In recent years, Magnetic Resonance Imaging (MRI) technology has emerged as a valuable tool in diagnosing AD due to its non-radiative, non-invasive, and non-harmful characteristics. In particular, it offers high tissue resolution and can be utilized for imaging with a variety of parameters (Jin et al., 2020; Cao et al., 2022). However, single modality-based investigations may not provide sufficient information to identify complex diseases. The multi-modal MRI techniques, such as diffusion MRI (dMRI) and functional MRI (fMRI), can provide a holistic view to observe changes in brain structure and function of AD (Cuingnet et al., 2011; Zhang et al., 2020). In the context, considering the advantages of complementary information, multi-modal analyses corresponding to both structural and functional characteristics have a great boom simultaneously (Poldrack and Farah, 2015). Furthermore, the brain network analysis has been widely employed in the diagnosis of brain diseases, which can provide valuable insights into the connected mechanisms between different brain regions (Lama and Kwon, 2021). For instance, dMRI (Soares et al., 2013) has been utilized to construct structural connectivity to measure the connections of nerve fiber bundles in white matter, while resting-state functional MRI (rsfMRI) (Sheline and Raichle, 2013; Soares et al., 2016) has been used to construct functional connectivity to detect the functional activity of the brain.

Recent advancements in AI, particularly Generative Adversarial Networks (GANs), have demonstrated great potential in analyzing complex brain data. GANs are capable of learning and generating new data samples that resemble the input data (Fahimi et al., 2020). In the context of AD, GANs can be trained on large datasets of brain images to learn patterns associated with the disease, helping in the diagnosis of AD by identifying subtle changes in brain structure or function. However, the current limitations of most AI models in neuroimaging classification tasks lead to underutilization of existing information and insufficient processing of unbalanced data. The primary challenge lies in the strategy of randomly selecting data for training at once, which ignores the potential benefits of utilizing data systematically and continuously.

Confronted with the complexity of these brain science problems, the Brain Informatics methodology provides a systematic perspective to understand the principles and mechanisms of human information processing related to high-order cognition functions cognitive functions (such as reasoning, calculation and problem solving) (Yang et al., 2009), as well as the development of new technologies for analyzing the biological characteristics and clinical applications on brain diseases. In the context of Brain Informatics, multi-modal and multi-scale brain data are analyzed systematically by considering different distributions of samples, so as to personalized applications. For example, the Data-Brain driven general intelligence model (Kuai and Zhong, 2020) is proposed to realize systematic brain computing in terms of the diversities of brain data from the experimental perspective. In particular, an iterated and evolved computing cycle was designed to continuously evidence combination and fusion computing.

In this paper, we propose the BNLoop-GAN model, which couples the Loop-based Generative Adversarial Network with the ECFC method for multiple loop brain network learning. The main contributions of this study can be summarized as follows: (1) an enhanced-GAN model is developed, utilizing techniques such as conditional generation, patch-based discrimination, and Wasserstein gradient penalty to learn the implicit distribution of brain networks; (2) a multiple-loop-learning algorithm is introduced, which combines evidence with better sample contribution ranking during continuous training phases; (3) the BNLoop-GAN model is applied to a case study of AD classification, where single-modal and multi-modal brain networks are computed iteratively to improve classification performance.

The rest of this paper is organized as follows: Section 2 provides a review of related works on brain networks, GANs, and AD. Section 3 introduces the overall framework of the BNLoop-GAN model for classification tasks, which comprises an enhanced GAN model and a multiple-loop-learning algorithm. Section 4 describes the experimental settings, data preparation, brain network construction, and performance evaluation. Section 5 presents results and discusses different scenarios on single-modal and multi-modal brain networks. Finally, Section 6 gives a conclusion and outlines future work.

2. Related work

Recently, AI models have gained widespread popularity in image generation, image super-resolution and other requirements based on their generative capabilities of addition, deletion, and modification. In the medical field, GAN models have been widely applied to diagnosis of AD. For instance, Yu et al. (2022) proposed a Multidirectional Perception GAN that uses a multidirectional mapping mechanism to learn morphological features for classifying AD severity at different stages. Yu et al. (2021) also proposed a three-player cooperative game-based framework with the high-order pooling scheme, namely tensorizing GAN, which is used to learn the structural information of MRI to assess mild cognitive impairment and AD. Moreover, a condition GAN (cGAN) model (Jung et al., 2023) is proposed to generate high-quality 3D MR brain images at different stages of AD, which integrates an additional module to ensure smooth and realistic transitions in 3D space, and uses an adaptive identity loss to preserve patient identification features. Ji et al. (2021) proposed a framework utilizing recurrent GANs for estimating effective connectivity from rsfMRI data, revealing potential differences in neural influence and information flow between AD and healthy control (HC) groups.

Given the complexity of AD, many studies have paid attention to use GAN models for multi-modal neuroimaging analysis. Pan et al. (2021) developed a Decoupling GAN to detect abnormal neural circuits for AD, which decomposes a brain network into two parts and utilizes an analytic module associated with the hyperedge neurons algorithm. The proposed model can extract complementary topology information between rsfMRI and diffusion tensor imaging (DTI) to detect abnormal neural circuits at different stages of AD. Moreover, a cross-modal transformer GAN (Pan and Wang, 2022) has been introduced, which employs a bi-attention mechanism to merge rsfMRI and DTI data effectively, facilitating the identification of AD-associated brain connectivity and enhancing the accuracy of classification. Zuo et al. (2021) developed a multi-modal representation learning and adversarial hypergraph fusion framework

using complete trimodal images (MRI, DTI and rsfMRI) to address the limitation of data distribution inconsistency in AD diagnosis. Zuo et al. (2021) also developed a prior guided adversarial representation learning and hypergraph perceptual network, which can evaluate the changing characteristics of brain connectivity at different stages of AD.

With the progress of brain connectivity, brain network analyses break a new ground in the study of AD. Cui et al. (2018) developed a minimum spanning tree method to construct the brain functional network, and extracted the topological features of the brain network. They used the support vector machine to compare AD and HC groups. Islam and Zhang (2018) proposed a deep convolutional neural network to learn features from a small and imbalanced dataset of structural MRI, which can identify and classify AD at different stages. Ye et al. (2019) selected DTI from 161 participants and used multivariate distance matrix regression (MDMR) analysis to detect structural abnormalities of brain networks during the development of AD disease. On the basis of the seed regions selected by MDMR analysis, supervised learning was applied to evaluate the predictive performance of AD. Furthermore, Zhang et al. (2022) proposed a multi-graph convolutional network based on GAN, which can learn the complex relationship between individual brain structural and functional networks automatically. Lei et al. (2021) proposed an automatic weighted centralized multi-task learning framework, in which multi-task learning is applied to identify features integration of structural and functional connectivity, for providing new insights into early AD detection.

Considering the complexity and systematization of brain computing in current big data era, the loop-based strategy is adopted to perform continuous learning inspired by Brain Informatics methodology. For example, Kuai et al. (2021) proposed the ECFC method to analyze multi-task fMRI data from different sources through merging systematic experimental design with evidence type reasoning. The uncertainty is analyzed and inferred to provide finer interpretations from both cognitive functions and brain regions. Furthermore, the similar strategy is adopted to decode the hidden relationship between connectivity abnormalities and brain disorders as well (Kuai et al., 2021). However, these methods only concern with fMRI at a single modal. In this paper, we extend the loop-inspired method from single modal to multiple modals, and from cognitive functions to brain diseases. In the next section, we will introduce how to realize a GAN-driven multiple-loop-learning to carry out systematic brain big data computing.

3. Methods

3.1. Overview

In this section, we introduce the overall framework for addressing the classification task of brain networks between abnormal and HC groups. The framework consists of three main components, which is illustrated in Figure 1. The first component is the brain network computing component, by which both structural and functional brain networks are obtained by analyzing multi-modal brain images. The second and third components are the enhanced-GAN model and the multiple-loop-learning algorithm, respectively, both of which constitute the Loop-based Generative Adversarial Network model for Brain Network (BNLoop-GAN). Before brain networks are learned, some preprocessing steps are required, including: the multi-modal MRI data are processed, such as denoising, calibration, correction; and then brain networks are constructed, such as brain region selection,

region segmentation, time series extraction and connectivity measures. Afterwards, the constructed brain networks are recognized by the BNLoop-GAN model with Classifiers to realize classification tasks. During this process, a multiple-loop-learning algorithm is used to select the small batch of samples from the whole training set step by step. The selected samples have an easier-to-learn probability distribution, which can reduce the complexity of model training. Each round of training processes is considered as a loop ($Loop - i, Loop - j | i, j \in N^+$), in which the same number of samples from abnormal and HC groups are selected ($S_{x,y,p,q,m,n} | x,y,p,q,m,n \in N^+$), where x,p,m represents the number of abnormal groups; y,q,n represents the number of HC groups; $x = y, p = q, m = n$. More specifically, a multiple-loop-learning algorithm is developed, depending on the training loop from previous iterations.

3.2. The enhanced-GAN model

To provide greater clarity on the enhanced-GAN model within the BNLoop-GAN model, illustrated in Figure 2, we present further details on its constituent components: a generator, a discriminator, and a classifier. The generator is structured with transposed convolutional layers, batch normalization, and activation functions such as ReLU and Sigmoid. Similarly, the discriminator and classifier share the similar structure, both of which are composed of convolutional layers, batch normalization, and LeakyReLU activation functions.

To provide support for the generation of matrices with specified attributes in the subsequent multiple-loop-learning algorithm, the enhanced-GAN model incorporates conditional information into both the generator and discriminator. This is accomplished through the use of a conditional GAN (cGAN) (Mirza and Osindero, 2014) architecture, which enables the model to better comprehend the contextual information of the generation task.

Patch-based processing is commonly utilized in computer vision tasks, including image analysis and object recognition, because it allows for local analysis of image features. This approach can be especially useful when dealing with complex symmetric matrices, as it enables the network to focus on smaller, more manageable sections of the input at a time. The idea of PatchGAN (Isola et al., 2017) is combined here to map the input to $N * N$ patches. These patches are designed to process matrices in a “patch-wise” manner, meaning that they divide the input image into small overlapping patches and process each patch individually. By learning the brain regions using block features, it is possible to gain a deeper understanding of how the network is processing and interpreting the input matrix at a local level. This information can be useful for identifying patterns or features within the brain regions that are important for the network’s decision-making process and for improving the performance of the network on the discrimination task.

Wasserstein GAN with gradient penalty (WGAN-GP) (Gulrajani et al., 2017) is added to address the problems of traditional GANs (Goodfellow et al., 2014), such as mode collapse and training instability. The core concept of WGAN-GP is to use Wasserstein distance to measure the difference between the generated and real data distributions and to enforce the Lipschitz continuity of the network through gradient penalty. Compared to traditional

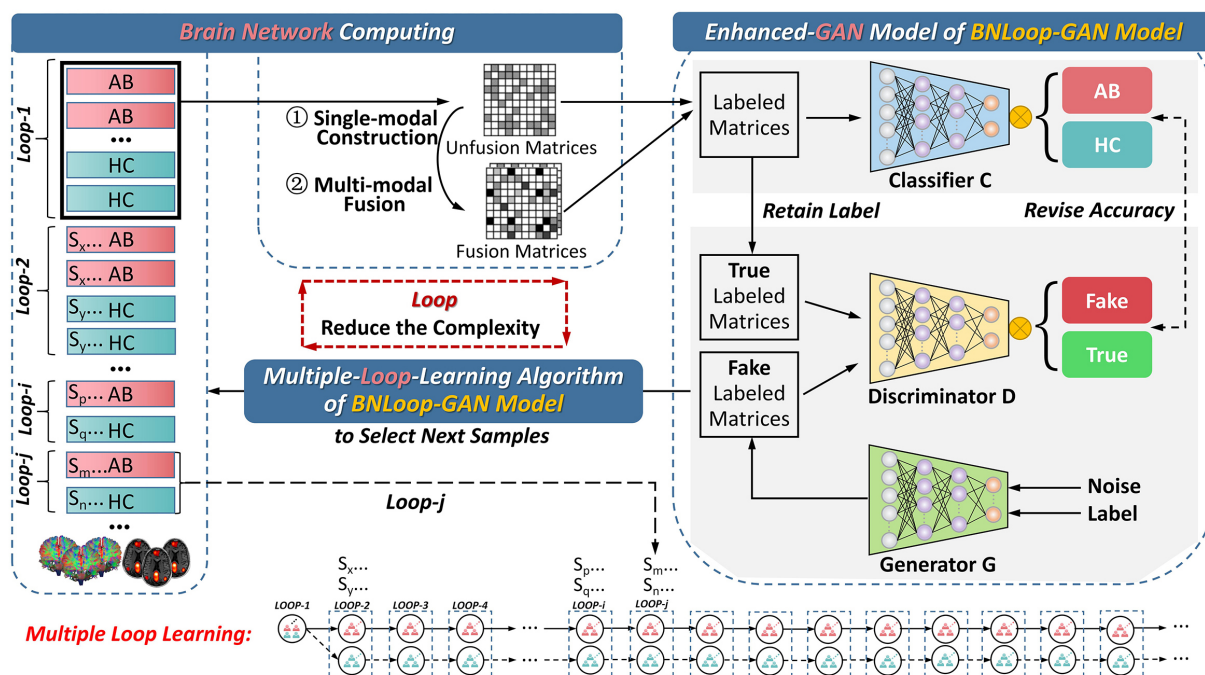


FIGURE 1

The overall framework for the classification of brain networks. AB, abnormal groups; HC, healthy control groups.

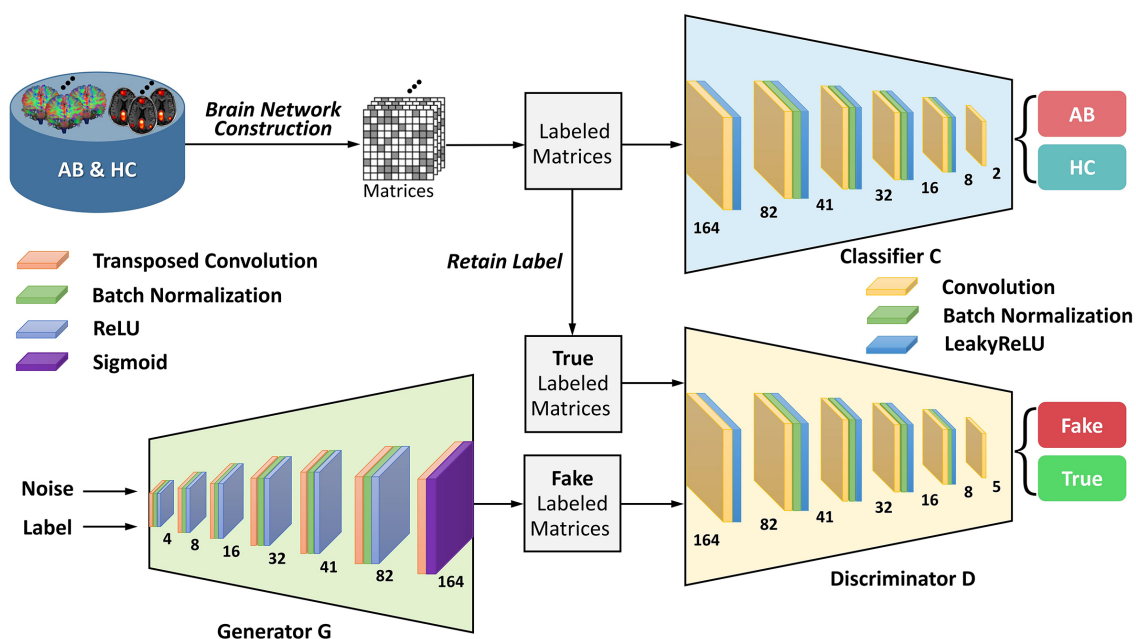


FIGURE 2

The architecture of the enhanced-GAN model. AB, abnormal groups; HC, healthy control groups.

Wasserstein GAN (Arjovsky et al., 2017), WGAN-GP has the advantage of providing more stable training performance and producing better sample quality of brain networks. The definition of Wasserstein distance is shown as follows:

$$W(P_r, P_g) = \inf_{\gamma \sim \Pi(P_r, P_g)} E_{(x,y) \sim \gamma} [\|x - y\|] \quad (1)$$

where P_r is the real distribution and P_g is the model distribution implicitly defined by the generator; $\Pi(p_r, p_g)$ denotes the set of all joint distributions $\gamma(x,y)$ whose marginal distributions are P_r and P_g respectively; $E_{(x,y) \sim \gamma} [\|x - y\|]$ is the mathematical expectation of distance $\|x - y\|$; and $\inf \{\cdot\}$ is the lower bound of set.

In order to solve the mode collapse and improve the convergence speed of traditional GANs, the gradient penalty is added to the

discriminator loss function, and the generated samples are constrained by Lipschitz. The discriminator loss function is:

$$L_D = E_{\tilde{x} \sim P_g} [D(\tilde{x})] - E_{x \sim P_r} [D(x)] + \lambda E_{\tilde{x} \sim P_g} \left[\left(\left\| \nabla_{\tilde{x}} D(\tilde{x}) \right\|_2 - 1 \right)^2 \right] \quad (2)$$

where λ represents the coefficient of the gradient penalty item; $P_{\tilde{x}}$ generates a straight-line uniform sampling between P_g and P_r ; $\nabla_{\tilde{x}} D(\tilde{x})$ is the gradient of the discriminator network; $\|\cdot\|_2$ stands for second norm of matrix.

3.3. Multiple-loop-learning algorithm

Contrary to most of existing methods that learn data using batch training, this framework splits whole dataset into different subsets for incremental learning. The flowchart of the multiple-loop-learning algorithm in Figure 3 illustrates how to generate incremental training plans.

As shown in Figure 3, in the first loop (*Loop* - i , $i=1$), the learning process starts by randomly selecting AB and HC samples with a predefined size from the database. Both the Generator and the Discriminator are pre-trained to drive the multiple-loop-learning algorithm. The pre-trained Generator is then employed to generate fake samples that are close to the true distribution of samples. The pre-trained Discriminator is used to process the real and generated samples to compute the sample contribution ranking using Euclidean Distance. To select samples with an easier-to-learn distribution during training loops, the top half of the ranking samples is chosen as the batch for training the classification model, and the middle half to three-quarters of the ranking is prepared for the next loop sample. The bottom quarter of the ranking is put back into the database. Thus far, the first round of loop ends when the classifier has learned a batch. The samples ranked in middle half to three-quarters of the previous round are combined with randomly selected samples to obtain AB and HC samples, which are used in a new round of loop. Multiple-loop-learning is achieved through continuous loop optimization, and the algorithm continues until the classifier converges. Algorithm 1 provides specific details of the algorithm.

Algorithm 1: BNLoop-GAN with multiple-loop-learning.	
Input:	the pre-training classifier C , generator G , and discriminator D ;
	the brain network matrix of healthy control groups, M_{HC} ;
	the brain network matrix of abnormal groups, M_{AB} ;
	the real matrix, RM ;
	the fake matrix, FM ;
	the table of sample contribution ranking of RM , T_{RM} ;
Output:	the trained classifier, C ;
	the accuracy of C , ACC ;
1: Initializing the loop = 0;	
2: Initializing the batch_samples;	
3: while C not converge	
4: random select M_{HC} and M_{AB} to fill $RM_{(M_{HC}, M_{AB})}^{loop}$ to batch_samples from the database;	
5: generate $FM_{(M_{HC}, M_{AB})}^{loop}$ from G of random noise;	
6: $T_{RM} \leftarrow$ Euclidean Distance computing $D(RM_{(M_{HC}, M_{AB})}^{loop})$ and $D(FM_{(M_{HC}, M_{AB})}^{loop})$;	
7: choose $RM_{(M_{HC}, M_{AB})}^{loop}$ of 1/2 top-ranked T_{RM} for training classifier C ;	
8: get ACC ;	
9: $RM_{(M_{HC}, M_{AB})}^{temp} \leftarrow RM_{(M_{HC}, M_{AB})}^{loop}$ of 1/2-3/4 top-ranked T_{RM} ;	
10: $RM_{(M_{HC}, M_{AB})}^{loop}$ of 1/4 bottom-ranked T_{RM} back to the database;	
11: loop + +;	
12: initializing $RM_{(M_{HC}, M_{AB})}^{loop}$;	
13: adding $RM_{(M_{HC}, M_{AB})}^{temp}$ to $RM_{(M_{HC}, M_{AB})}^{loop}$;	
14: end while	
15: return C , ACC	

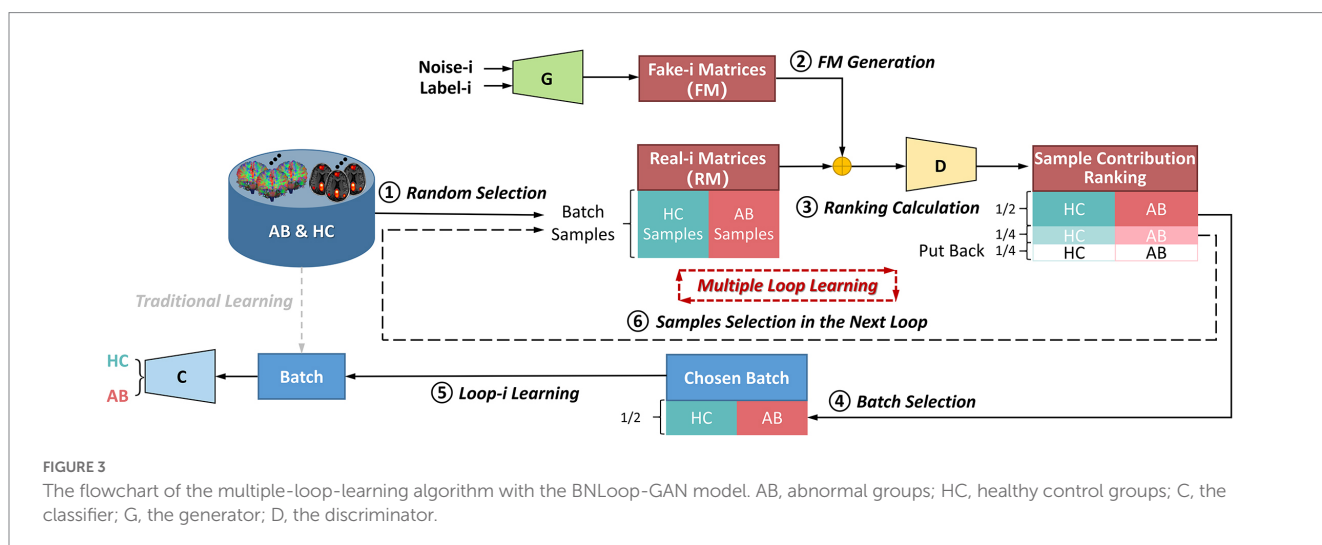


FIGURE 3

The flowchart of the multiple-loop-learning algorithm with the BNLoop-GAN model. AB, abnormal groups; HC, healthy control groups; C, the classifier; G, the generator; D, the discriminator.

4. Experiments

4.1. Datasets and preprocessing

In this paper, the MRI data (including dMRI and rsfMRI) were gathered from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.¹ These subjects were instructed to rest with their eyes open, not to think of anything in particular, and not to fall asleep while collecting rsfMRI. The data set contains 42 AD patients (72.0 ± 17.0, 30F/93M) and 42 gender-age matched HC groups (74.5 ± 10.9, 39F/92M).

To start with, the raw MRI data were converted from DICOM to NIfTI using “dcm2nii” function in the MRICroGL software.² The bvec and bval files were generated to calculate various diffusion properties on the diffusion gradients and directions. All diffusion-based tractography approaches and subsequent connectome reconstructions were performed in the MRtrix3 software.³ Firstly, the initial diffusion images were denoised to increase signal-to-noise ratio. Secondly, gibbs-ringing and bias field correction were performed to reduce artifacts and non-reguliarities. The eddy current-induced distortion was removed, and head motion error was corrected. Finally, the mean b0 image generated by averaging all the images with b = 0 s/mm² was used to register the diffusion image to the structural MRI using the FSL toolbox.⁴ The rsfMRI data were processed by SPM12 software⁵ with the standard procedures, including slice-timing correction, realignment to the median image, and co-registration to the individual structural MRI.

4.2. Brain network computing

4.2.1. Construction of diffusion MRI networks

The constrained spherical deconvolution (CSD) method overcomes the limitations of crossing fibers inherent in the diffusion tensor model (Tournier et al., 2008). Therefore, we performed multi-shell multi-tissue CSD method to obtain the fiber orientation distribution (FOD) (Jeurissen et al., 2014). The white matter pathways of whole brain were reconstructed using probabilistic streamline tractography through the second-order integration over FOD algorithm (Smith et al., 2013). The aparc2009 template (Destrieux et al., 2010) of FreeSurfer was used to divide each brain region, and the connection strength is normalized by the number of streamlines divided by the brain volume, thereby constructing structural brain networks.

4.2.2. Construction of resting-state functional MRI networks

The functional brain networks were constructed using the Nilearn package in Python.⁶ For each subject, the average time series of each

brain region were extracted using the aparc2009 template of FreeSurfer. Then, the connectivity characteristics were measured using the Pearson correlation coefficient as shown in Equation 3, by which the matrices of N*N-dimensional functional connectivity were obtained for each subject.

$$r(X, Y) = \frac{\sum_{i=1}^N [X(i) - \bar{X}] * [Y(i) - \bar{Y}]}{\sqrt{\sum_{i=1}^N [X(i) - \bar{X}]^2 * \sum_{i=1}^N [Y(i) - \bar{Y}]^2}} \quad (3)$$

where $r(X, Y)$ is the Pearson correlation coefficient to measure connected effects between brain regions X and Y ; $X(i)$ and $Y(i)$ represent the time series from two different brain regions respectively, $i = 1, 2, \dots, N$, and N is the number of time points of the subject; \bar{X} and \bar{Y} are the mean values of $X(i)$ and $Y(i)$ respectively.

4.2.3. Joint learning of multi-modal brain networks

Brain disorders exhibit multi-aspect changes in the brain's structural, functional and dynamic characteristics frequently. The structure forms the foundation of function, while the function is the representation of structure. Multi-modal MRI data analyses can capture complementary characteristics from diverse perspectives, bringing richer information and benefiting classification tasks consequently. We adopted the joint learning method of multi-modal data, that is, superimposing the number of dMRI and rsfMRI brain network channels. The follow-up experimental results can reflect its advantages compared with single-modal brain networks.

4.2.4. Brain network augmentation

In order to reduce noise and facilitating normalization of input features, we performed min-max scaling, which involves scaling the data to a range between 0 and 1. Additionally, to prevent over-fitting, the data augmentation techniques were used to expand the training data set, as shown in Figure 4.

For a given image in Figure 4, it can be seen that an original matrix (A) is transformed by moving its first column to the last column, generating an in-process matrix (B), and then its first row is moved to the last row, resulting in the enhanced matrix (C). In this way, the strategy of data augmentation will avoid breaking the symmetry of the matrix. We repeat this process on the newly generated enhanced matrix (C), generating 163 additional enhanced matrices from one original brain network matrix corresponding to a single subject, and 13,692 enhanced matrices from 84 subjects in total. We employed all of these matrices, with 80% reserved for training and 20% for testing.

4.3. Model description and evaluation indicators

Table 1 presents the architectural parameters of the BNLoop-GAN model in detail. This model is capable of accommodating both single-modal and multi-modal inputs, with the parameter ‘ n ’ in Table 1 denoting the number of modalities.

¹ <https://adni.loni.usc.edu>

² <https://www.nitrc.org/projects/mricrogl>

³ <https://www.mrtrix.org>

⁴ <https://www.fmrib.ox.ac.uk/fsl>

⁵ <https://www.fil.ion.ucl.ac.uk/spm/software/spm12>

⁶ <https://nilearn.github.io>

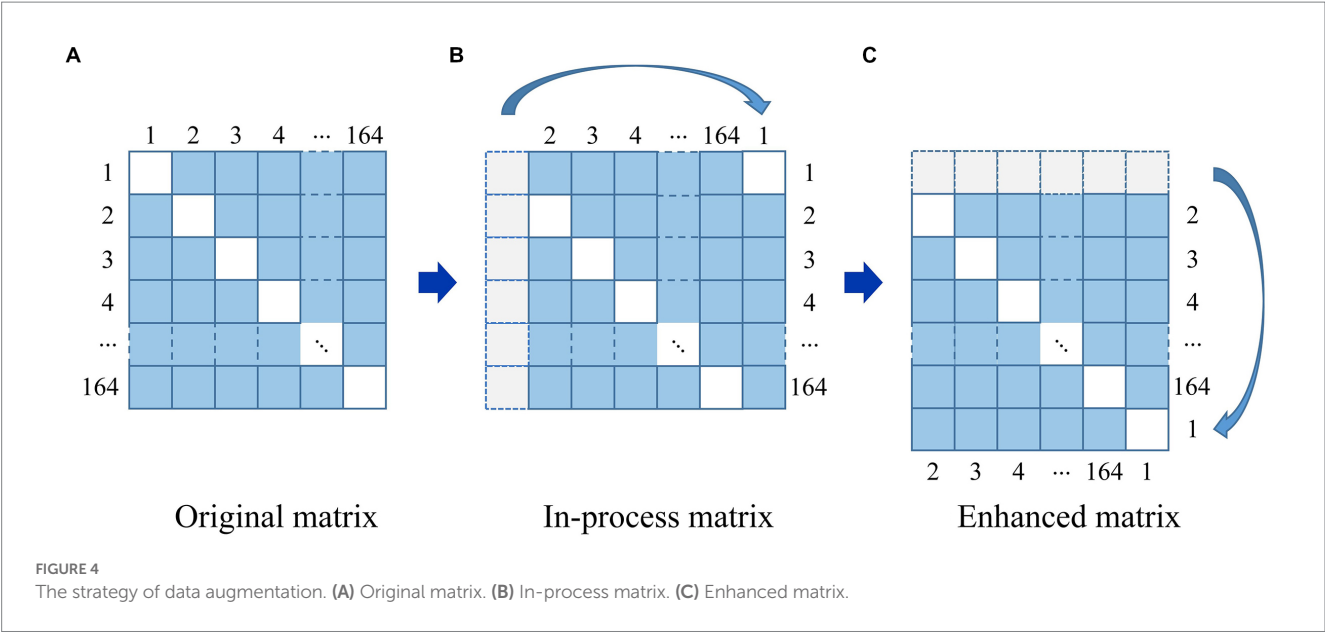


TABLE 1 The architectural parameters of the BNLoop-GAN model.

Sub-module	Layer	Kernel	Stride	Padding	Size of feature map (height × width × channels)	Activation function
Generator	Input	–	–	–	1 × 100 × <i>n</i>	–
	ConvT 1	4	1	0	4 × 4 × 512	ReLU
	ConvT 2	4	2	1	8 × 8 × 256	ReLU
	ConvT 3	4	2	1	16 × 16 × 128	ReLU
	ConvT 4	4	2	1	32 × 32 × 64	ReLU
	ConvT 5	12	1	1	41 × 41 × 32	ReLU
	ConvT 6	4	2	1	82 × 82 × 16	ReLU
	ConvT 7	4	2	1	164 × 164 × <i>n</i>	Sigmoid
Discriminator	Input	–	–	–	164 × 164 × <i>n</i>	–
	Conv 1	4	2	1	82 × 82 × 16	LeakyReLU
	Conv 2	4	2	1	41 × 41 × 64	LeakyReLU
	Conv 3	10	1	0	32 × 32 × 128	LeakyReLU
	Conv 4	4	2	1	16 × 16 × 256	LeakyReLU
	Conv 5	4	2	1	8 × 8 × 512	LeakyReLU
	Conv 6	4	1	0	5 × 5 × <i>n</i>	–
Classifier	Input	–	–	–	164 × 164 × <i>n</i>	–
	Conv 7	4	2	1	82 × 82 × 16	LeakyReLU
	Conv 8	4	2	1	41 × 41 × 64	LeakyReLU
	Conv 9	10	1	0	32 × 32 × 128	LeakyReLU
	Conv 10	4	2	1	16 × 16 × 256	LeakyReLU
	Conv 11	4	2	1	8 × 8 × 512	LeakyReLU
	Conv 12	8	2	1	2 × 2 × 1	–

ConvT, Transposed Convolution; Conv, Convolution.

Three indicators are used to evaluate the performance of the model, including accuracy (*ACC*), sensitivity (*SEN*), specificity (*SPE*). The formula is defined as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

(4)

$$SEN = \frac{TP}{TP + FN}$$

(5)

$$SPE = \frac{TN}{FP + TN}$$

(6)

where FP, FN, TP, and TN denote False Positive, False Negative, True Positive and True Negative assessments, respectively.

Furthermore, due to the complexity of AD diagnosis, the Receiver Operating Characteristic (ROC) curve and Area under the ROC Curve (AUC) are utilized to evaluate the efficacy of binary classification models. The ROC curve plots the true positive rate (*TPR*) against the false positive rate (*FPR*), with *TPR* on the y-axis and *FPR* on the x-axis.

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

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

The AUC is the area under the ROC curve, with values ranging from 0 to 1. A higher AUC indicates better model performance, with 0.5 indicating random guessing and 1 indicating perfect prediction.

5. Results and discussions

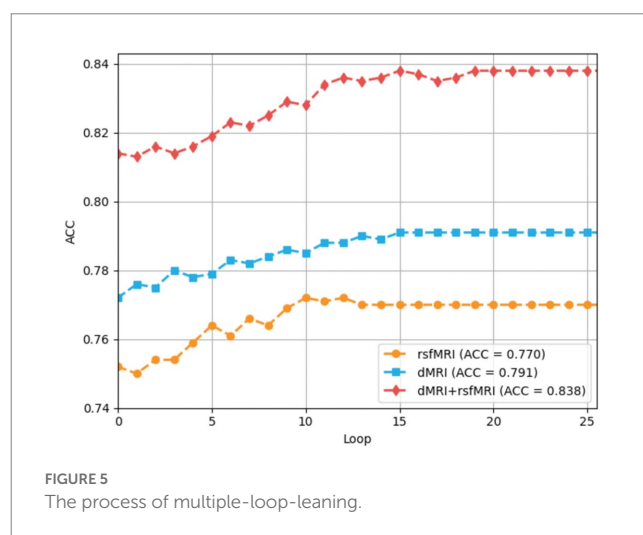
A case study was performed to examine the classification performance corresponding to single-modal brain networks and multi-modal brain networks, respectively, through the BNLoop-GAN model with the multiple-loop-learning algorithm. In order to evaluate the effectiveness of the different components incorporated into the enhanced-GAN model, we conducted a series of ablation experiments. Table 2 presents the results of these experiments, which were evaluated using three indicators.

Taking into account the necessity of conditional information for driving multiple-loop-learning algorithms, we conducted a series of Loop-based ablation experiments on the baseline FCN model. Compared to the baseline FCN, the FCN with loop of conditional generation demonstrated no significant improvement in indicators for any modality. Simply capturing the similarity of samples through conditional information is not enough to improve loop efficiency. However, FCN with loop of conditional generation and patch-based discrimination, as well as FCN with loop of conditional generation and Wasserstein gradient penalty, both improved performance of the classifier for all modalities. The patch-based discrimination ensures that the generated images have a high degree of similarity to real images in terms of brain regions. The Wasserstein gradient penalty enforces the Lipschitz continuity constraint in the discriminator. Both of them contributes to the improved quality of generated samples. It is worth noting that the BNLoop-GAN model exhibited the highest performance for all three modalities, combining techniques of conditional generation, patch-based discrimination, and Wasserstein gradient penalty to learn the implicit distribution of brain regions. These techniques optimize the model by improving the quality of generated samples, selecting samples with an easier-to-learn distribution during training loops, and providing better performance on a classification task of brain networks. In addition, the evaluation indicators for multi-modal data are higher than those for single-modal data.

TABLE 2 The results of AD classification of different brain network learning strategies using various models.

Model	Brain Network Learning Strategy	ACC (%)	SEN (%)	SPE (%)
FCN	①	77.2	74.7	79.6
	②	75.3	73.9	80.3
	③	81.4	80.2	82.2
FCN with loop of conditional generation	①	77.3	74.5	79.8
	②	75.1	73.6	79.9
	③	81.6	80.5	82.5
FCN with loop of conditional generation and patch-based discrimination	①	78.2	75.1	80.5
	②	76.1	74.2	81.2
	③	82.3	81.0	83.1
FCN with loop of conditional generation and Wasserstein gradient penalty	①	77.8	75.2	80.6
	②	75.9	74.1	81.1
	③	82.1	80.9	83.0
BNLoop-GAN	①	79.1	76.3	81.2
	②	77.0	75.2	81.9
	③	83.8	81.8	84.9

FCN is the fully convolutional neural network. Three brain network learning strategies were utilized, including: ① Single-modal brain networks based on dMRI; ② Single-modal brain networks based on rsfMRI; ③ Multi-modal brain networks based on dMRI and rsfMRI. The bold values represent the highest accuracy, sensitivity and specificity of the BNLoop-GAN model on Multi-modal brain networks based on dMRI and rsfMRI.



The training processes of each loop driven by the multiple-loop-learning algorithm in the BNLoop-GAN model are shown in Figure 5. The overall trend of the training process reveals that the model's accuracy can be improved steadily and effectively, regardless of whether single-modal or multi-modal data is used. Furthermore, it can be seen that multi-modal brain networks learned by the BNLoop-GAN model achieve the better accuracy of 83.8% than others related to single-modal brain networks.

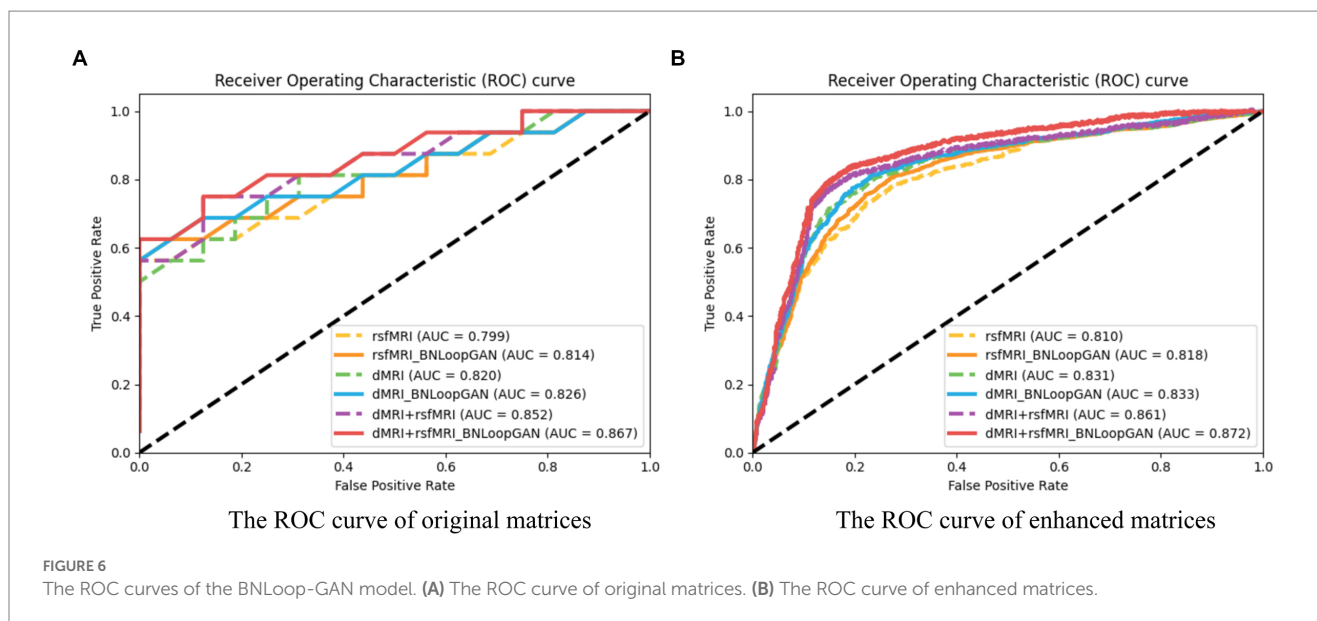


Figure 6 illustrates the computation of ROC curves, which provide a comprehensive representation of performance across different brain network types. Sub-figure (A) and (B) are constructed using two strategies of original matrices (without brain network augmentation) and enhanced matrices (with brain network augmentation) as a test set to verify the effectiveness of the model. The experimental results prove that data augmentation actually increase the performance of the model. The six curves of each sub-figure are represented by different colors corresponding to different modalities and different model strategies. It can be clearly seen that the classification effect of the model based on multi-modal data is significantly higher than that of single-modal data. Additionally, each dotted ROC curves represent the performance of the basic mode (i.e. FCN), whereas the solid ROC curves depict the performance of the entire BNLoop-GAN model using the multiple-loop-learning algorithm. Compared with the AUC value of the single classifier model, the classification of BNLoop-GAN model has a slight improvement, indicating that the effectiveness of the multiple-loop-learning algorithm can improve the performance of classification, and it performs better in the use of multi-modal data. The AUC value reaches 0.872. All experiments are performed in the same experimental environment with the parameters of the equipment (Intel(R) Core(TM) CPU i7-8750H @ 2.20GHz, 12 CPU cores, 8GB NVIDIA GeForce GTX 1070).

In recent years, an increasing number of studies utilize multiple modalities, such as dMRI and fMRI, to improve the classification of brain networks. Various combination techniques, including feature selection (Yu et al., 2022), data augmentation (Venugopalan et al., 2021), transfer learning (Ghaffari et al., 2022), and more, have been proposed to optimize classification results. For instance, Meng et al. (2022) proposed the multi-modal LassoNet model, which combines fMRI and DTI modalities in a sparse Lasso neural network framework and incorporates connection strength and subject structure to construct a comprehensive multi-modal brain network. The model has achieved a classification accuracy of approximately 90.68% for AD-HC. Mohtasib et al. (2022) conducted a comprehensive connectivity analysis between the default mode

network regions using group independent component analysis on rsfMRI data, and examined the paired structural connectivity between the frontal lobe region and the hippocampus using DTI data. They applied both logistic regression and random forest models to classify AD patients and HC groups, achieving an accuracy of 74%. Although some current studies can achieve higher accuracy, it is worth noting that the evaluation strategy is based on k-fold cross-validation (Alorf and Khan, 2022) which is difficult to transfer into real-world scenarios. In this paper, we consider the test samples are not seen in the training phrase. The advantages of the proposed BNLoop-GAN model are as follows. Firstly, an enhanced-GAN model is designed for facilitating to learn the implicit distribution of the brain networks. Secondly, it utilizes the multiple-loop-learning algorithm to select easier-to-learn samples during training loops, continuously improving model classification performance. Lastly, the model can achieve satisfy performance on classification tasks of AD using multi-modal brain network fusion.

6. Conclusion

In this paper, the BNLoop-GAN model with a multiple-loop-learning algorithm is proposed to the classification of brain diseases from the brain network perspective. The proposed model is evaluated by the AD classification task, using rsfMRI, dMRI, and their fusion. The experimental results show that the fused brain image learning can achieve a better performance than others, strengthening the importance of fusing structural and functional information. Moreover, the loop learning mode can effectively learn the implicit distribution of brain networks to reduce training complexity and improve classification performance. In the future, more effort will be required to solve the following issues, such as: expanding multi-modal MRI data such as task-state fMRI to capture deeper feature patterns; designing the reasoning rules for representing the main and supplementary modal types with weights and their relations; enriching the “evidence combination-fusion computing” methods for multi-modal brain 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 authors.

Author contributions

YC: conceptualization of this study, methodology design and implementation, analysis, interpretation of data, and writing original draft. HK: conceptualization of this study, methodology design and implementation, analysis, and interpretation of data. PL: interpretation of data. J-SP: interpretation of data. NZ: conceptualization of this study, methodology design, interpretation of data, and final approval of the version. JY: conceptualization of this study, interpretation of data, and final approval of the version. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors thank the reviewers and editors for their constructive comments and guidance.

References

- Alorfi, A., and Khan, M. U. G. (2022). Multi-label classification of Alzheimer's disease stages from resting-state fMRI-based correlation connectivity data and deep learning. *Comput. Biol. Med.* 151:106240. doi: 10.1016/j.combiomed.2022.106240
- Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein GAN. *ArXiv*, Preprint. doi: 10.48550/arXiv.1701.07875
- Cao, Y., Kuai, H., and Peng, G. (2022). "Enhancing the MR neuroimaging by using the deep super-resolution reconstruction. Brain informatics" in *15th International Conference, BI 2022, Padua, Italy, July 15–17, 2022, Proceedings* (Cham: Springer International Publishing), 184–194.
- Citron, M. (2010). Alzheimer's disease: strategies for disease modification. *Nat. Rev. Drug Discov.* 9, 387–398. doi: 10.1038/nrd2896
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M. O., et al. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *NeuroImage* 56, 766–781. doi: 10.1016/j.neuroimage.2010.06.013
- Cui, X., Xiang, J., Guo, H., Yin, G., Zhang, H., Lan, F., et al. (2018). Classification of Alzheimer's disease, mild cognitive impairment, and Normal controls with subnetwork selection and graph kernel principal component analysis based on minimum spanning tree brain functional network. *Front. Comput. Neurosci.* 12:31. doi: 10.1148/radiol.10100734
- Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic Parcellation of human cortical Gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010
- Fahimi, F., Dosen, S., Ang, K. K., Mrachacz-Kersting, N., and Guan, C. (2020). Generative adversarial networks-based data augmentation for brain-computer Interface. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 4039–4051. doi: 10.1109/TNNLS.2020.3016666
- Ferrari, C., and Sorbi, S. (2021). The complexity of Alzheimer's disease: an evolving puzzle. *Physiol. Rev.* 101, 1047–1081. doi: 10.1152/physrev.00015.2020
- Ghaffari, H., Tavakoli, H., and Pirzad, J. G. (2022). Deep transfer learning-based fully automated detection and classification of Alzheimer's disease on brain MRI. *Br. J. Radiol.* 95:20211253. doi: 10.1259/bjr.20211253
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. *arXiv*, Preprint. doi: 10.48550/arXiv.1406.2661
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A., et al. (2017). Improved training of Wasserstein Gans. *Adv. Neural Inf. Proces. Syst.* 30:5327. doi: 10.5555/3295222.3295327
- Islam, J., and Zhang, Y. (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an Ensemble system of deep convolutional neural networks. *Brain Informat.* 5:2. doi: 10.1186/s40708-018-0080-3
- Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A. (2017). "Image-to-Image Translation with Conditional Adversarial Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 5967–5976.
- Jeurissen, B., Tournier, J. D., Dhollander, T., Connelly, A., and Sijbers, J. (2014). Multi-tissue constrained spherical Deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage* 103, 411–426. doi: 10.1016/j.neuroimage.2014.07.061
- Ji, J., Liu, J., Han, L., and Wang, F. (2021). Estimating effective connectivity by recurrent generative adversarial networks. *IEEE Trans. Med. Imaging* 40, 3326–3336. doi: 10.1109/TMI.2021.3083984
- Jin, D., Wang, P., Zalesky, A., Liu, B., Song, C., Wang, D., et al. (2020). Grab-AD: generalizability and reproducibility of altered brain activity and diagnostic classification in Alzheimer's disease. *Hum. Brain Mapp.* 41, 3379–3391. doi: 10.1002/hbm.25023
- Jung, E., Luna, M., and Park, S. H. (2023). Conditional GAN with 3D discriminator for MRI generation of Alzheimer's disease progression. *Pattern Recogn.* 133:109061. doi: 10.1016/j.patcog.2022.109061
- Kuai, H., Chen, J., Tao, X., Imamura, K., Liang, P., and Zhong, N. (2021). "Exploring the brain information processing mechanisms from functional connectivity to translational applications" in *Brain informatics: 14th international conference, BI 2021, virtual event, September 17–19, 2021, proceedings 14* (New York: Springer International Publishing), 99–111.
- Kuai, H., Tao, X., and Zhong, N. (2022). Web intelligence meets brain informatics: towards the future of Artificial intelligence in the connected world. *World Wide Web* 25, 1223–1241. doi: 10.1007/s11280-022-01030-5
- Kuai, H., and Zhong, N. (2020). The extensible data-brain model: architecture, applications and directions. *J. Comput. Sci.* 46:101103. doi: 10.1016/122j.jocs.2020
- Kuai, H., Zhong, N., Chen, J., Yang, Y., Zhang, X., Liang, P., et al. (2021). Multi-source brain computing with systematic fusion for smart health. *Informat. Fusion* 75, 150–167. doi: 10.1016/j.inffus.2021.03.009
- Lama, R. K., and Kwon, G. R. (2021). Diagnosis of Alzheimer's disease using brain network. *Front. Neurosci.* 15:605115. doi: 10.3389/fnins.2021.605115
- Lei, B., Cheng, N., Frangi, A. F., Wei, Y., Yu, B., Liang, L., et al. (2021). Auto-weighted centralised multi-task learning via integrating functional and structural connectivity for subjective cognitive decline diagnosis. *Med. Image Anal.* 74:102248. doi: 10.1016/j.media.2021.102248
- Meng, X., Liu, J., Fan, X., Bian, C., Wei, Q., Wang, Z., et al. (2022). Multi-modal neuroimaging neural network-based feature detection for diagnosis of Alzheimer's disease. *Front. Aging Neurosci.* 14:451. doi: 10.3389/fnagi.2022.911220

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2023.1202382/full#supplementary-material>

- Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *ArXiv Preprint*. doi: 10.48550/arXiv.1411.1784
- Mohtasib, R., Alghamdi, J., Jobeir, A., Masawi, A., Pedrosa de Barros, N., Billiet, T., et al. (2022). MRI biomarkers for Alzheimer's disease: the impact of functional connectivity in the default mode network and structural connectivity between lobes on diagnostic accuracy. *Heliyon* 8:e08901. doi: 10.1016/j.heliyon.2022.e08901
- Pan, J., Lei, B., Wang, S., et al. (2021). DecGAN: Decoupling generative adversarial network detecting abnormal neural circuits for Alzheimer's disease. *ArXiv Preprint*. doi: 10.48550/arXiv.2110.05712
- Pan, J., and Wang, S. (2022). Cross-modal transformer GAN: A brain structure-function deep fusing framework for Alzheimer's disease. *ArXiv Preprint*. doi: 10.48550/arXiv.2206.13393
- Poldrack, R. A., and Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature* 526, 371–379. doi: 10.1038/nature15692
- Sheline, Y. I., and Raichle, M. E. (2013). Resting state functional connectivity in preclinical Alzheimer's disease. *Biol. Psychiatry* 74, 340–347. doi: 10.1016/j.biopsych.2012.11.028
- Smith, R. E., Tournier, J. D., Calamante, F., and Connelly, A. (2013). SIFT: spherical-deconvolution informed filtering of Tractograms. *NeuroImage* 67, 298–312. doi: 10.1016/j.neuroimage.2012.11.049
- Soares, J. M., Magalhães, R., Moreira, P. S., Sousa, A., Ganz, E., Sampaio, A., et al. (2016). A Hitchhiker's guide to functional magnetic resonance imaging. *Front. Neurosci.* 10:515. doi: 10.3389/fnins.2016.00515
- Soares, J. M., Marques, P., Alves, V., and Sousa, N. (2013). A Hitchhiker's guide to diffusion tensor imaging. *Front. Neurosci.* 7:31. doi: 10.3389/fnins.2013.00031
- Tournier, J. D., Yeh, C. H., Calamante, F., Cho, K. H., Connelly, A., and Lin, C. P. (2008). Resolving crossing Fibres using constrained spherical Deconvolution: validation using diffusion-weighted imaging phantom data. *NeuroImage* 42, 617–625. doi: 10.1016/j.neuroimage.2008.05.002
- Venugopalan, J., Tong, L., Hassanzadeh, H. R., and Wang, M. D. (2021). Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci. Rep.* 11:3254. doi: 10.1038/s41598-020-74399-w
- Yang, Y., Liang, P., Lu, S., Li, K., and Zhong, N. (2009). The role of the DLPFC in inductive reasoning of MCI patients and normal agings: An fMRI study. *Sci. China Series C: Life Sci.* 52, 789–795. doi: 10.1007/s11427-009-0089-1
- Ye, C., Mori, S., Chan, P., and Ma, T. (2019). Connectome-wide network analysis of white matter connectivity in Alzheimer's disease. *NeuroImage Clin.* 22:101690. doi: 10.1016/j.nicl.2019.101690
- Yu, J., Kong, Z., Zhan, L., Shen, L., and He, L. (2022). Tensor-based multi-modality feature selection and regression for Alzheimer's disease diagnosis. *ArXiv Preprint*. doi: 10.48550/arXiv.2209.11372
- Yu, W., Lei, B., Ng, M. K., Cheung, A. C., Shen, Y., and Wang, S. (2021). Tensorizing GAN with high-order pooling for Alzheimer's disease assessment. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 4945–4959. doi: 10.1109/TNNLS.2021.3063516
- Yu, W., Lei, B., Wang, S., Liu, Y., Feng, Z., Hu, Y., et al. (2022). Morphological feature visualization of Alzheimer's disease via multidirectional perception GAN. *IEEE Trans. Neural Netw. Learn. Syst.* PP, 1–15. doi: 10.1109/TNNLS.2021.3118369
- Zhang, L., Wang, L., and Zhu, D. Alzheimer's Disease Neuroimaging Initiative (2022). Predicting brain structural network using functional connectivity. *Med. Image Anal.* 79:102463. doi: 10.1016/j.media.2022.102463
- Zhang, Y. D., Dong, Z., Wang, S. H., Yu, X., Yao, X., Zhou, Q., et al. (2020). Advances in multi-modal data fusion in neuroimaging: overview, challenges, and novel orientation. *Informat. Fusion* 64, 149–187. doi: 10.1016/j.inffus.2020.07.006
- Zhong, N., Bradshaw, J. M., Liu, J., and Taylor, J. G. (2011). Brain informatics. *IEEE Intell. Syst.* 26, 16–21. doi: 10.1109/MIS.2011.83
- Zhong, N., Hu, J., Motomura, S., Wu, J.-L., and Liu, C. (2005). Building a data-mining Grid for multiple human brain data analysis. *Comput. Intell.* 21, 177–196. doi: 10.1111/j.0824-7935.2005.00270.x
- Zuo, Q., Lei, B., Shen, Y., Liu, Y., Feng, Z., and Wang, S. (2021). "Multimodal representations learning and adversarial Hypergraph fusion for early Alzheimer's disease prediction" in *Pattern recognition and computer vision: 4th Chinese Conference, PRCV 2021, Beijing, China, October 29–November 1, 2021, Proceedings, Part III 4* (New York: Springer International Publishing), 479–490.
- Zuo, Q., Lei, B., Wang, S., et al. (2021). A prior guided adversarial representation learning and Hypergraph perceptual network for predicting abnormal connections of Alzheimer's disease. *ArXiv Preprint*. doi: 10.48550/arXiv.2110.09302



OPEN ACCESS

EDITED BY

Yong Hu,
The University of Hong Kong, Hong Kong SAR,
China

REVIEWED BY

Liya Ding,
Southeast University, China
Jing Teng,
North China Electric Power University, China

*CORRESPONDENCE

Camille Simon-Chane
✉ camille.simon-chane@ensea.fr

RECEIVED 10 May 2023

ACCEPTED 24 July 2023

PUBLISHED 15 August 2023

CITATION

Brémond-Martin C, Simon-Chane C,
Clouchoux C and Histace A (2023) Brain
organoid data synthesis and evaluation.
Front. Neurosci. 17:1220172.
doi: 10.3389/fnins.2023.1220172

COPYRIGHT

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

Brain organoid data synthesis and evaluation

Clara Brémond-Martin^{1,2}, Camille Simon-Chane^{1*},
Cédric Clouchoux² and Aymeric Histace¹

¹ETIS Laboratory UMR 8051 (CY Cergy Paris Université, ENSEA, CNRS), Cergy, France, ²Witsee, Neoxia,
Paris, France

Introduction: Datasets containing only few images are common in the biomedical field. This poses a global challenge for the development of robust deep-learning analysis tools, which require a large number of images. Generative Adversarial Networks (GANs) are an increasingly used solution to expand small datasets, specifically in the biomedical domain. However, the validation of synthetic images by metrics is still controversial and psychovisual evaluations are time consuming.

Methods: We augment a small brain organoid bright-field database of 40 images using several GAN optimizations. We compare these synthetic images to the original dataset using similitude metrics and we perform an psychovisual evaluation of the 240 images generated. Eight biological experts labeled the full dataset (280 images) as synthetic or natural using a custom-built software. We calculate the error rate per loss optimization as well as the hesitation time. We then compare these results to those provided by the similarity metrics. We test the psychovalidated images in a training step of a segmentation task.

Results and discussion: Generated images are considered as natural as the original dataset, with no increase of the hesitation time by experts. Experts are particularly misled by perceptual and Wasserstein loss optimization. These optimizations render the most qualitative and similar images according to metrics to the original dataset. We do not observe a strong correlation but links between some metrics and psychovisual decision according to the kind of generation. Particular Blur metric combinations could maybe replace the psychovisual evaluation. Segmentation task which use the most psychovalidated images are the most accurate.

KEYWORDS

psychovisual, metric, validation, brain organoid, AAE

1. Introduction

The scarcity of public datasets of annotated biomedical images remains an unresolved bottleneck to develop specialized and robust analysis tools. Often, research groups do not share experimental data for privacy reasons. The high costs of equipment, long acquisition times, and necessary in-depth expertise can be a brake to acquisitions by other teams (Chakradhar, 2016). To benefit from the advances in deep-learning (DL) for automated image analysis, large training datasets are necessary. Moreover, original dataset constraints create a problem of class imbalance with deep learning training procedures. These problems are emphasized with small sets, reduced to a few images (Tajbakhsh et al., 2016).

Data augmentation is widely used in the biomedical domain to increase the size of image datasets (Singh and Raza, 2021). While classical data augmentation, based on transformations (flip-flops, rotation, whitening, etc.), does not increase the diversity of the dataset, a solution widely provided in the biomedical field is the use of Generative Adversarial Networks (GAN) which produce new synthetic images from natural ones (Yi et al., 2019; Lan et al., 2020).

GANs are unsupervised deep-based architectures composed of generator and a discriminator. The generator aims at creating visually realistic and natural images while the discriminator tries to decipher whether the result is generated (Goodfellow et al., 2014). Both networks are trained simultaneously with the same loss function. Since its first creation, multiple GAN architecture variations have been proposed to generate and extend biomedical datasets (Lan et al., 2020; Fernandez et al., 2021; Chen et al., 2022). However, the validation of these synthetic images remains a challenge Alqahtani et al. (2019). Two main evaluating methods are existing the non-automated based upon psychophysics methods in perceptual psychology, and automated methods based upon metrics (Salimans et al., 2016a; Zhou et al., 2019). Psychovisual evaluation is a time consuming gold standard which requires many subjects to reduce its intrinsic subjectivity. On the other hand, there is no commonly approved specific metric to evaluate whether GAN-generated synthetic images can be considered as natural. The use of common metrics is also controversial (Borji, 2019).

We consider the case of brain organoids, three dimensional cultures differentiated from pluripotent stem cells (Lancaster et al., 2013). After a neural induction, brain organoids present cell communications, organized tissues and an organization similar to the brain with various regions such as neuro-epithelial zones (Kelava and Lancaster, 2016). They complement *in vivo* brain models to follow physiological and pathological brain development. However, brain organoids suffer from batch syndrome: in the same culture environment they do not innately develop comparable morphologies (Lancaster et al., 2013).

There are no specific tools to study the development of brain organoids. Though it may seem natural to implement machine learning algorithms to aid this task, few brain organoid image datasets are publicly available. Between January 2018 and June 2020 for instance only six over 457 articles in the brain organoid field let the image datasets in openaccess and only one concerns brain organoids in bright-field (Brémond Martin et al., 2021). The emergence of brain organoids has created a new field, few laboratories have the knowledge and experience necessary to grow these cultures. Pandemic restrictions have further limited the possibility for several teams to grow and image such culture over the past few years.

The largest public brain-organoid image dataset we know of contains 40 images (Gomez-Giro et al., 2019). Data augmentation solutions have already been used to increase the size and diversity of this brain organoid bright-field dataset. An adversarial autoencoder (AAE) seems the architecture the most suited to augment images of brain organoid bright-field acquisition (Brémond Martin et al., 2022a). This AAE differs from the original GAN architecture by the input given to the encoding part (original images) and its generative network containing an auto-encoder-decoder framework (Goodfellow et al., 2014; Makhzani et al., 2016). The encoder learns to convert the data distribution to the prior distribution, while the decoder learns a deep generative model that maps the imposed prior to the data distribution thanks to a latent space (Makhzani et al., 2016). As is typical with AAEs, the images generated are visibly blurry.

To improve the sharpness during the generation, we test various loss functions to improve the adversarial network

(Brémond Martin et al., 2022a). However, these results are based upon metric calculation and a dimensional reduction to compare all feature images (original and generated with each optimization) in the same statistical space. Indeed some metrics may not be suited to identify the naturality of an image as they are originally created to test the similitude or the quality of images (Borji, 2019; Brémond Martin et al., 2022a). In our previous contribution we observe the data augmentation strategy based upon AAE loss optimizations used during the training step of a DL segmentation algorithm improve the quality of the shape extraction of brain organoids (Brémond Martin et al., 2022a). The first raised question is does these images seem as natural as the original images to furnish a better segmentation quality compared to a result issue from classic data augmentation strategies? Another fundamental issue remains unresolved: do these synthetic images seem natural to a biological expert point of view as for metric(s)? Psychovisual evaluations have been already made on others bright-field cell synthetic cell generation (Malm et al., 2015). This evaluation is an important step for the validation of a particular generative model of images. Thus the selected images as natural by Human Biological experts could maybe help to train deep based segmentation methods and characterize their development but with now a double psychovisual-metric validation.

We propose to evaluate the synthetic images generated by an AAE (Brémond Martin et al., 2022a) using both with similarity metrics and biological experts. The purpose of this article is to give the lacking non-automated psychovisual evaluation of the synthetic images which is a new contribution compared to our previous contribution which focus on automated metric based and statistical strategies in order to understand if the naturality of these images could explain the segmentation results (Brémond Martin et al., 2022a). The second original part of the work is to find a metric combination which may replace or complete the psychovisual non-automated evaluation. Related work is presented in the following section. Section 3 describes the generative network, the metric evaluation and the psychovisual evaluation. Section 4 successively shows the results for the metric evaluation and the psychovisual evaluation, followed by a cross-analysis of the two evaluation methods. These results are analyzed in the Section 5. Section 6 sums up the main contributions of this paper.

2. Related work

The first generative adversarial network, proposed by Goodfellow et al. (2014), is constituted by two connected networks: the generative model (*GM*) maps the images into the space (*z*) by an objective function (*F*); the discriminative model (*DM*) determines the probability for which a point from *z* belongs to the original dataset (*o*) or to the generated dataset (*g*). Training *F* increases the probability that the data synthesized is attributed to *o*. The probability of correct sample labeling (belonging to generated *g* or original *o*) is maximized by *D*. Simultaneously, *GM* is trained to leverage the discriminator function expressed by:

$$\min_{GM} \max_{DM} F(DM, GM) = E_o p_{data}[\log D_o] + E_g p_g[\log(1 - D(G_z))]. \quad (1)$$

An overview of various GAN architectures used for medical imaging is given by Yi et al. (2019). Previous work compares five GAN architectures commonly used for biological datasets to find the best suited network to increase a small database of bright-field images of organoids (Brémond Martin et al., 2022a): the original GAN implementation (Goodfellow et al., 2014); CGAN gives to the generator input the correct label (physiological or pathological) (Mahmood et al., 2018); DCGAN by Radford et al. (2015) is constituted by a convolutional neural networks instead of the generator; INFOGAN uses the generated images at an epoch to train the subsequent (Hu et al., 2019); AAE by Makhzani et al. (2016) uses an auto-encoder as a generator. This work evaluates the generated images using metrics.

As already reported in the literature by Lan et al. (2020), GAN and CGAN produce mode collapse with such a small dataset. In addition, GAN produces a white imprint around the shape of the organoid. DCGAN and INFOGAN generate a divergent background which makes the images difficult to exploit. This is probably due to the high variability of the small dataset, as mentioned in Lan et al. (2020). Only AAE produces images of similar quality to the original and seems the best architecture to generate images with a few input. However, AAE generates blurry images and the background differs from the light-to-black gradient present in many bright-field images.

Noise-injection during the AAE generation was studied to improve the background generation with satisfying results in Brémond Martin et al. (2022b). In this paper, we further explore the effect of loss optimization on the AAE architecture to improve the sharpness of the generated images. The results are evaluated by an automatic approach based on metrics and by a psychovisual study. The last part aims at giving an application of such psychovalidated images in a segmentation task.

3. Methods

In this section, we first present the generative methods and optimizations used. We follow by a description of the metric and psychovisual evaluations including the experimental setup, datasets and experts. We present the analysis methodologies of the neurobiologist decisions and the comparison of psychovisual evaluations with metric calculations. We finish by giving an asset of the importance of psychovalidated images in an augmented dataset strategy for a segmentation task.

3.1. Generative adversarial networks

3.1.1. Original images

Our dataset is composed of 40 microscope acquisitions and 240 synthetic images created by a AAE loss optimizations. “Original” images are the 40 bright-field brain organoid acquisitions provided by Gomez-Giro et al. (2019). This dataset is made of image of 20 physiological and 20 pathological organoids acquired over three days on the same apparatus. These input images ($1,088 \times 1,388$ pixels) are cropped and scaled to 250×250 pixels, maintaining their original proportions. A scale factor of 4 is chosen so the scripts can

run in a reasonable amount of time without downgrading the input image quality too much.

3.1.2. Loss optimizations

The images generated by the AAE architecture we use are somewhat blurry. To overcome this phenomenon we study how the discriminator loss can influence the quality of the image generation. We consider six losses: the Binary Cross Entropy (BCE) which is most commonly used in GANs and five other losses which are specifically known to improve the contrast or sharpness of the generated images.

BCE is the most commonly used loss for GANs and the baseline of this work. It is calculated by:

$$\text{BCE} = -\frac{1}{n} \sum_{i=1}^n (y_i(\log(y'_i)) - ((1 - y_i)(\log(1 - y'_i))) \quad (2)$$

with y the real image tensor and y' the predicted ones (Makhzani et al., 2016) and n the number of training.

Summing the L1 norm to the BCE is reported to reduce overfitting (Wagnier-Dauchelle et al., 2019). We hypothesize this norm could improve the quality of the generation as reported in image restoration tasks which does not over-penalize large errors (Zhao et al., 2017).

$$\text{L1} = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i|_1 \quad (3)$$

$$\text{BCE}_{\text{L1}} = \text{BCE} + \alpha \text{L1}. \quad (4)$$

α is set to 10^{-4} , as in the original paper.

The least square loss (LS) is reported by Mao et al. (2017) to avoid gradient vanishing in the learning process step resulting in better quality images:

$$\text{LS} = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (5)$$

A Poisson loss is used in Wagnier-Dauchelle et al. (2019) to improve the sensitivity of a segmentation task:

$$L_{\text{Poisson}} = \frac{1}{n} \sum_{i=1}^n (y'_i - y_i) \log(y'_i + \epsilon) \quad (6)$$

where ϵ is a regularization term set to 0.25.

The DeblurGAN was developed to unblur images using the Wasserstein loss (Kupyn et al., 2018). Since we are also interested in deblurring the output images, we have tested this loss with the proposed AAE where $(P(y, y'))$ is the joint distributions of y and y' for which the distributions are equal to P_y and $P_{y'}$, and $p(y, y')$ the proportion of y or y' to move to have $P_y = P_{y'}$:

$$Wass(P(y, y')) = \sum_{i=1}^n \inf_{p(P(y, y'))} E_{y_i, y'_i} \delta p(||y_i - y'_i||) \quad (7)$$

However, we do not apply a l2 content loss such as in Kupyn et al. (2018) added to the Wasserstein loss, or add a penalty gradient

to the Wasserstein loss such as in [Gulrajani et al. \(2017\)](#). Our dataset containing various subclasses (physiological and pathological brain organoid images acquired at three developmental stages), we aim at creating a Normalized Wasserstein loss to avoid imbalanced mixture proportions ([Balaji et al., 2019](#)). We apply the L2 normalization on the Wasserstein loss, producing a new loss we call the Perceptual Wasserstein loss for the first time to our knowledge, applied on an AAE architecture:

$$P.Wass(P(y, y')) = \sum_{i=1}^n \sqrt{\inf_{p(y_i, y'_i)} E_{y_i, y'_i} \delta p(\|y_i - y'_i\|)^2} \quad (8)$$

3.1.3. Training

[Figure 1](#) shows the global training setup. The 40 original images are used to generate 40 synthetic images for each architecture (each loss). Input and output images measure 250×250 pixels. Training lasts 2,000 epochs for each optimization; this corresponds to the plateau before over-fitting for all loss optimizations.

3.1.4. Resources

The GAN algorithms are developed in Python 3.6 with an Anaconda framework containing the 2.3.1 Keras and 2.1 Tensorflow versions. All scripts are executed on an Intel Core i7-9850H CPU with 2.60 GHz and a NVIDIA Quadro RTX 3000s GPU device.

3.2. Metric evaluation

Six metrics are used to compare the similitude of the synthetic images generated by the AAE to the original dataset. A blur metric is used to evaluate the quality of these synthetic images.

The Frechet Inception Distance (FID) is calculated between two groups of images ([Heusel et al., 2017](#)). This score tends toward low values when the two groups (original O or generated G images) are similar, with μ the average value of the pixels of all images of a group, and Σ the covariance matrix of a group:

$$FID(O, G) = \|\mu_O - \mu_G\|^2 + \text{Tr}(\Sigma_O + \Sigma_G - 2(\Sigma_O \Sigma_G)^{\frac{1}{2}}) \quad (9)$$

The Structural Similarity Index (SSIM) is calculated using luminance, contrast and structure by [Wang et al. \(2004\)](#) between two images o and g belonging respectively to O and G .

$$SSIM(o, g) = \frac{(2\mu_o\mu_g + c_1)(2\sigma_{og} + c_2)}{(\mu_o^2 + \mu_g^2 + c_1)(\sigma_o^2 + \sigma_g^2 + c_2)} \quad (10)$$

where σ represents the standard deviation, c_1 is a constant that ensures the luminance ratio is always positive when the denominator is equal to 0, and c_2 is an other constant for the contrast stability. The SSIM ranges between 0 (no similitude) and 1 (high similitude).

The Universal Quality Metric (UQM) is based on the calculation of the same parameters as SSIM and was proposed by

[Wang and Bovik \(2002\)](#). UQM ranges between 0 and 1 (1 being the highest quality):

$$UQM(o, g) = \frac{4\mu_o\mu_g\mu_{og}}{(\mu_o^2 + \mu_g^2)(\sigma_o^2 + \sigma_g^2)} \quad (11)$$

Entropy-based Mutual Information (MI) measures the correlation between original and generated images and ranges between 0 (no correlation) and 1 (high correlation) ([Pluim et al., 2003](#)):

$$MI(o, g) = \sum_{o \in O} \sum_{g \in G} P(o, g) \log \frac{P(o, g)}{P(o)P(g)} \quad (12)$$

where $P(o, g)$ is the joint distribution of o belonging to O and g from G .

The Mean Square Error (MSE) between an original image and a synthetic image is calculated as:

$$MSE(o, g) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (o(i, j) - g(i, j))^2 \quad (13)$$

The Peak Signal to Noise Ratio (PSNR) indicates a high signal power against noise, as used in [Jiang et al. \(2021\)](#). High values correspond to qualitative images. Pixels in images are ranked between 0 and 255, thus the maximum pixel value of an image is noted $\max(o)$ and equals at most 255.

$$PSNR(o, g) = 20 \log \max(o) - 20 \log MSE(o, g) \quad (14)$$

where \log denotes the common logarithm.

As a quality metric we calculate the blur index proposed by [Tsomko et al. \(2008\)](#) based on local image variance. A low score corresponds to a sharp image. In the following equation, the image is of size (m, n) , the predictive residues a given image pixel are $p(i, j)$ and their median $p'(i, j)$:

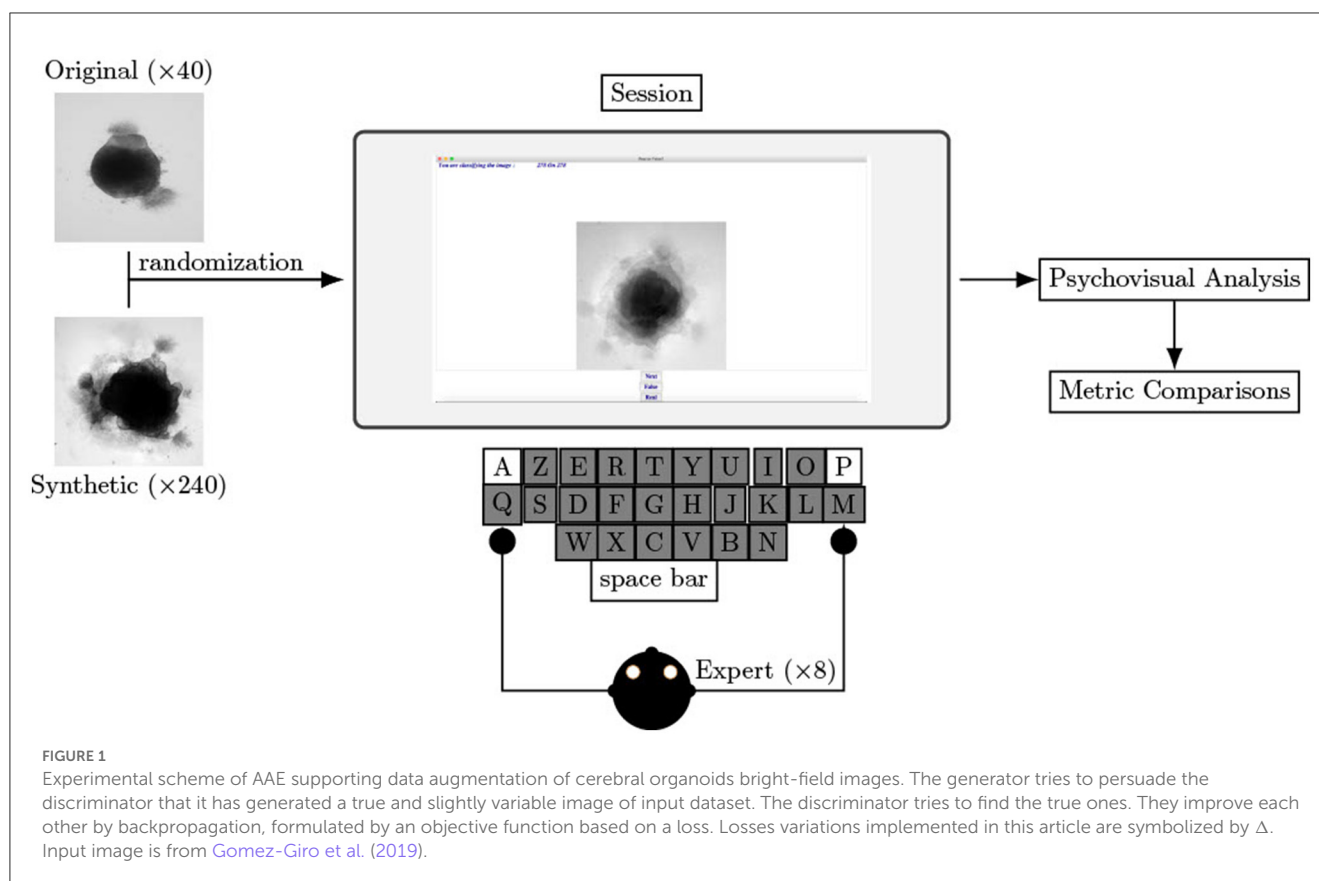
$$\text{Blur} = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^{n-1} [p(i, j) - p'(i, j)]^2 \quad (15)$$

The FID is designed to compare groups of images. We thus successively compare each group of synthetic images with the original input images. The FID reference range is calculated on the original image developmental stages. The SSIM, UQM, PSNR, MI, and MSE are designed to compare two images. For each group of synthetic images (all six losses) we successively compare every image with each original image and then compute the average of these 40×40 values. We also calculate these values on all pairs of original images to compare the results to the original range. The Blur index is calculated on individual images. We store the minimum and maximum value of this index for the original images and the average value per loss for the synthetic images.

3.3. Psychovisual evaluation

3.3.1. Dataset of original and synthetic images

The dataset to evaluate contains two classes of images the 40 microscopic acquisitions mentionned in Section 3.1.1 and the



240 synthetic images created by the previous mentioned AAE loss optimizations: binary cross entropy (BCE), binary cross entropy with a L1 normalization (BCE + L1), least squares (LS), Poisson, Wasserstein (Wass.), perceptual Wasserstein (P. Wass). The size of these 280 images is 250×250 pixels.

3.3.2. Randomization

To perform a double blind test the images are not labeled during the visualization so neither the experts or the team can be biased by the images information (real, generated, nor its kind of generation). Real and generated images are randomized at each test run. Each biological expert evaluates the complete randomized dataset (280 images). The randomization and corresponding labels are stored in a .csv file which is only accessible for result analysis.

3.3.3. Experts

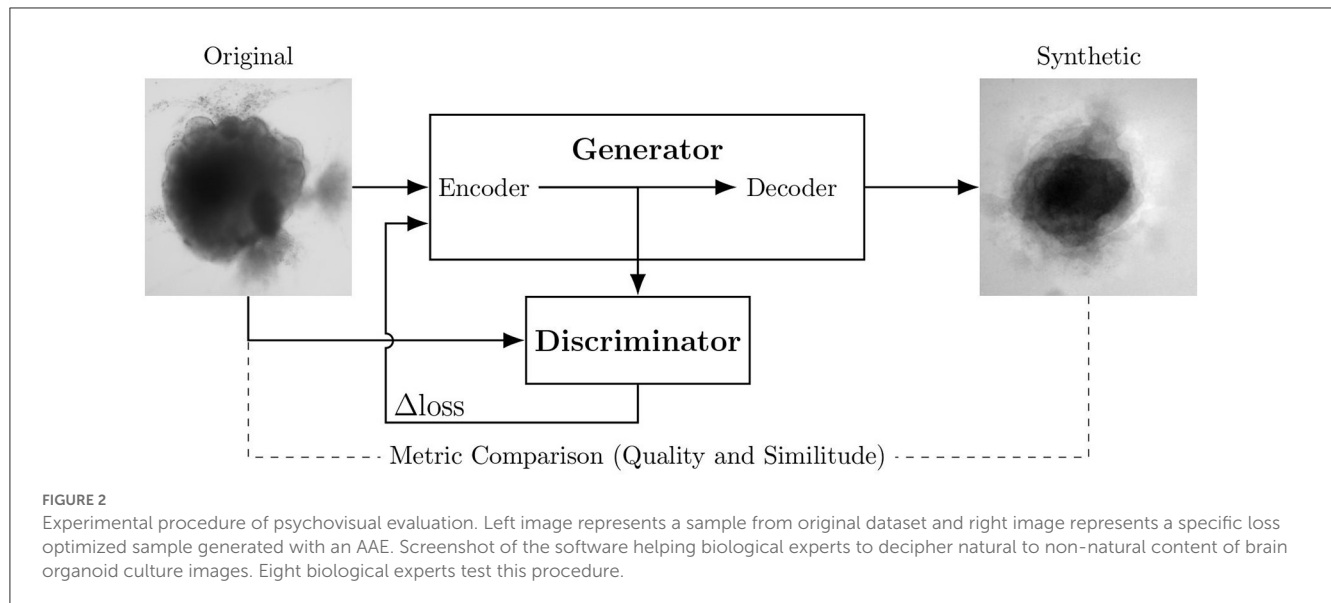
The experts who evaluated the database are biologists from ERRMECe laboratory, EA1391, CY Cergy Paris University. The group of eight experts is composed of three men and five women who are either PhD students, research engineers or researchers. They all have an expertise in neuronal culture and microscopy acquisition. We do not allow duplicate evaluators across evaluation procedure. During each evaluation session the evaluator is physically isolated from the other participants, without knowledge of other experts responses or images labels.

3.3.4. Evaluation software

To help experts in their evaluation and to ensure consistency throughout the entire experiment, we built a dedicated software using Python 3.6, as shown [Figure 2](#). The interface consists of an image displayed (250×250 pixels), a cursor with three buttons: *Real*, *Generated*, *Next* and the number of remaining images to classify. Keyboard shortcuts are available for *Real*, *Generated*, and *Next* buttons (respectively A, P, and Tab keys on an AZERTY keyboard) to facilitate the process. Images on the screen are updated each time *Next* is hit (or the corresponding keyboard shortcut). Clicks on *Next* are counted as a pass if not preceded by a *Real* or *Generated* click.

3.3.5. Experimental protocol

An operator enters the expert name and the date and hour of the recording. All eight experts chose to use their own mouse with the experiment laptop. The protocol, consisting of a single session and including all 280 images (real and synthetic), is described in [Figure 2](#). A pass is considered as an answer. The decision and answer time are saved at each click in a .csv file only accessible to analyze the results. The operator is present nearby to verify the smooth functioning of the experimental process and to capture any comments made by the experts. The list of questions the expert has to answer after the process is listed below:



- Why do you classify this generated image as a false? The operator shows the generated image with the longest hesitation.
- Why do you classify this original image as a false? The operator shows the original image with the longest hesitation, if this situation exists.
- What should we improve in future sessions?

We summarize the answers to these questions in the results section.

Each evaluation run produces two .csv files: One stores the randomization i.e. the order and label of each image presented. The second stores the experts' name and for each image present the answer time and the decision.

3.4. Analysis

The first analysis step consists in associating the randomization and the results files. We obtain, for each expert and for each image the decision and the decision time. The decision is then labeled as true positive (TP) or false negative (FN) for the original images and false positive (FP) or true negative (TN) for the synthetic images.

3.4.1. Parameter calculation

It is then standard to calculate the error rate (ER), defined as the number of false decisions divided by the total number of decisions.

$$ER = \frac{FP + FN}{FP + FN + TP + TN} \quad (16)$$

For the original images, this becomes:

$$ER_O = \frac{FN}{FN + TP} \quad (17)$$

and for the synthetic images:

$$ER_G = \frac{FP}{FP + TN}. \quad (18)$$

However, we wish to compare the proportion of synthetic images falsely labeled as true, with the proportion of original images label as true. We thus calculate the Positive Rate (PR) of the original images:

$$PR_O = 1 - ER_O = \frac{TP}{FN + TP}. \quad (19)$$

For the synthetic images $PR_G = ER_G$.

As a second parameter, we calculate the decision occurrence for each modality for each subgroup by a simple counting and render it in a % according to the total effective of a group of images.

We also evaluate the number of positive answers given by each expert as a count and the number of images given as a positive by zero expert, one expert, two experts etc. or the eight experts. Time decision and all these parameters are calculated between original and generated images, or between original and each modality of loss generation (BCE, BCE + L1, LS, Poisson, Wass., P. Wass.), globally or by each decision subgroups (FP, FN, TP, TN). All results are rendered as bargraphs representing variables (Time Decision in seconds or occurrence in % or error rates) according to one or many factors (group and subgroups of decision).

3.4.2. Metrics vs. human decision

To verify if some metrics highlight the same loss as producing the most natural images as experts, we plot each metric values for each loss group by each decision factor modality (FP, FN, TP, TN). In the dot representations for each loss group, each metric is plotted according to the Normalized Error Rate NER with individuals decision time t for the FP and for FN modality:

$$NER = \frac{FP \times t_{FP} + FN \times t_{FN}}{FP + FN + TP + TN} \quad (20)$$

where t_{FP} and t_{FN} are the average decision time respectively for FP and FN. In practice, this becomes, for the original images:

$$NER_O = \frac{FN \times t_{FN}}{FN + TP} \quad (21)$$

and for the synthetic images:

$$NER_G = \frac{FP \times t_{FP}}{FP + TN}. \quad (22)$$

To verify if a relation exists between a metric or a particular combination of metrics (M) and the decision or the time decision (DT), we calculate KL divergences on dimensional reduction results (Joyce, 2011).

$$KL = \sum_{x \in X} M_x \times \log \left(\frac{M_x}{DT_x} \right) \quad (23)$$

We only represent here kl-divergence corrpplots for each individual metric (and not metric combinations) for space considerations. We consider all possible metric combinations C :

$$C = \left\{ \binom{n}{k} \mid k \in N^*, k \leq n \right\} \quad (24)$$

where $n = 6$ is the total number of metrics. The total number of metric combinations considered is thus 63.

We then calculate Pearson and Kendall correlations between metric combinations and KL-div results (for error rate and time decision) for original and synthetic groups. We show the Pearson correlation for the ten best metric combinations. The best representation of these results (error rate or time by KL divergence) are represented as a scatter plot.

3.4.3. Statistical analysis

The normality is verified by a Shapiro and quantile to quantile graphics. We verify the homocedasticity in normal cases by a Bartlett test and in the case of non-normality by a Levene test. In case of a normality and homocedasticity, arametric tests are used (Anova), and non-parametrics tests otherwise (Kruskall–Wallis with a Tuckey *post hoc* test). Regression models are implemented to verify the interaction of factors (group and decision) on a specific variable (time or error rate or occurrence for instance). We use a *post-hoc* Holm test to compare two by two the effect of factors on variables after the regression.

We take an alpha risk at 5%. Correlation matrices are based on Pearson correlation tests.

3.5. Segmentation task

3.5.1. Dataset

We first build a training dataset composed the 40 images from the original dataset and 40 images obtained by flip-flops, rotations, whitenings, or crops of these original images. This “classical” dataset composes our baseline. We build five “psychovisual” training datasets where we replace part of the classically augmented images by synthetic images which are validated by 0, 2, 4, 6, or 8 experts.

All datasets are composed of 80 images of which 40 original but the proportion of synthetic images decreases as the number of experts required to validate an image increases. Ground truth has been manually segmented with the ITK-SNAP software (Yushkevich et al., 2006).

3.5.2. U-Net

Segmentation allows the extraction of an image content from its background. Various segmentation procedures exist but we have chosen U-Net which is widely used in the biomedical field (Ronneberger et al., 2015). U-Net has the advantage of working well for small training sets with data augmentation strategies, and has already been used for the ventricle segmentation of cleared brain organoids (Albanese et al., 2020).

3.5.3. Training

To robustly evaluate the performance of the segmentation a these small datasets we use a leave-one-out strategy where we only test on the original images. This results in 40 training sessions per dataset. We stop the training at 1,000 epochs with an average time of training of more than 1 h for each leave-one-out loop (six cases of augmentations \times 40 images = 240 h almost for the total training step). The summary of the leave-one-out strategy for every tested case is summarized in Figure 3.

3.5.4. Comparison of segmentations

To compare ground truth cerebral organoid content segmentation (GT) and U-Net (u) ones in various conditions, mean Dice scores are calculated as:

$$Dice_{(GT,u)} = \frac{2|GT \cap u|}{|GT| + |u|} \quad (25)$$

Thanks to the TP , FP , TN , and FN we could calculate the Accuracy, the Specificity, the Sensitivity, and the F1-score. The Accuracy is the ratio of true on the positives labels:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (26)$$

The Sensitivity is the ratio between how much were correctly identified as positive to how much were actually positive:

$$Sensitivity = \frac{TP}{TP + FN} \quad (27)$$

The Specificity is the ratio between how much were correctly identified as negative to how much were actually negatives:

$$Specificity = \frac{TN}{TN + FP} \quad (28)$$

The Precision is the ratio between how much were correctly identified as positives to how much were actually labeled as positives:

$$Precision = \frac{TP}{TP + FP} \quad (29)$$

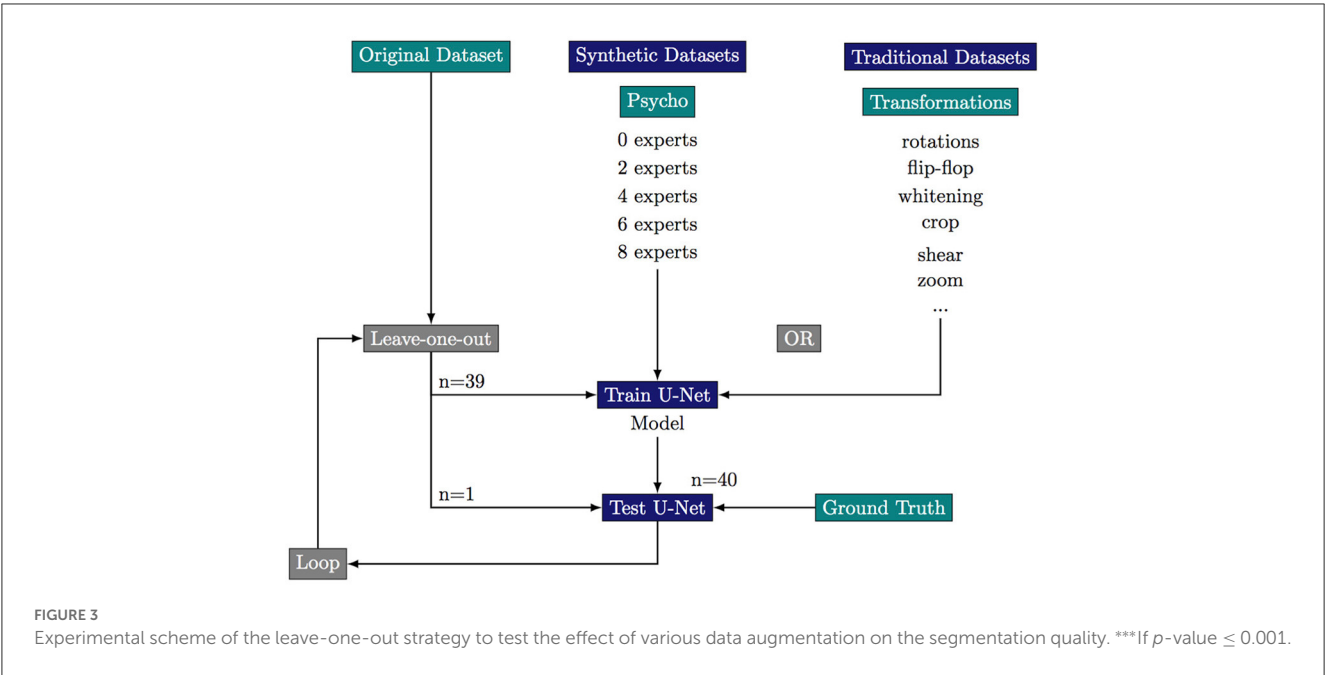
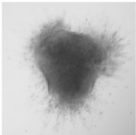
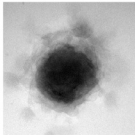
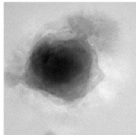
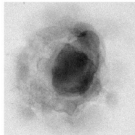
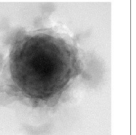
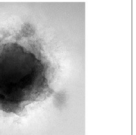
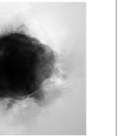
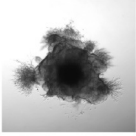
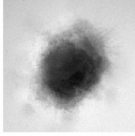
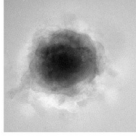
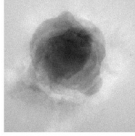
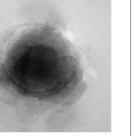
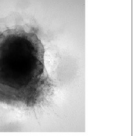
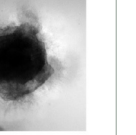
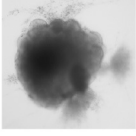
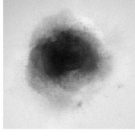
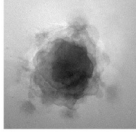
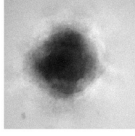
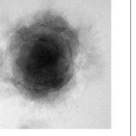
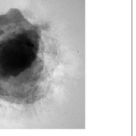



TABLE 1 Samples of original images and synthetic images generated by the AAE.

Original	BCE	BCE + L1	LS	Poisson	Wasserstein	P. Wass.
						
						
						

We show 3 of the 40 images for each group: original and per AAE loss variation.

The F1-Score allow to summarize the precision and the recall (Sensitivity) in an unique metric:

$$F1 - Score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity}$$

(30)

3.5.5. Visualization

To highlight real/false positive/negative segmentation we create a superimposed image composed by the ground truth and a sample of each segmentation resulting from the various trainings. We update the pixels values in lightpink the *FP* cerebral organoid segmentations and, in lightgreen the *FN*.

4. Results

We first present the metric evaluation of the synthetic images, then the results of the psychovisual evaluation of these synthetic images, and finally the correlations between the metrics and the psychovisual evaluation.

4.1. Qualitative evaluation

Table 1 shows three sample input images and three of the 40 synthetic images generated by each of the six AAE variations. Some of the generated samples are blurry and present a white imprint (BCE, BCE + L1, LS). Others show sharper edges and less visible

TABLE 2 Metric evaluation of AAEGAN brain organoid bright-field generated images.

Metric	Best	Original	BCE	BCE + L1	LS	Poisson	Wass.	P. Wass.
FID	Low	0.47–0.80	1.20	1.41	1.33	1.41	1.10	0.82
SSIM	High	0.65–0.71	0.63	0.62	0.60	0.63	0.62	0.50
UQM	High	0.63–0.87	0.83	0.83	0.84	0.84	0.83	0.82
MI	High	0.21–0.47	0.37	0.39	0.36	0.41	0.46	0.42
Blur	low	0.10–86.28	135.93	116.30	135.01	106.71	59.84	59.00
PSNR	Low	11.9–16.6	13.47	13.74	13.53	13.74	13.17	12.86
MSE	Low	93.25–106.23	103.13	103.35	104.01	103.33	103.11	102.93

We have calculated metrics on generated images from each AAE loss variations, with the BCE loss as the baseline. Scores outside the original range are in gray, best values are displayed in bold.

imprints (Poisson, Wasserstein and P. Wass.). For this group of three losses, only a few of the generated images seem to be identical to a given input image. For example the Poisson loss produces three images which are a blurred version of the original. These networks do not suffer from mode collapse.

4.2. Metric evaluation

To quantitatively confirm the visual analysis of the generated images, we calculate several metrics on both the original and synthetic images. These results are summarized in Table 2. This table presents the range of values given by the original images and the average value of each group of synthetic images (per loss optimization).

All six groups of synthetic images are within range of the original for the UQM, MI, PSNR, and MSE metrics. Images generated with a Poisson or LS loss have the highest UQM index. MI and MSE reach the best scores for the generated images using the Wasserstein losses. The FID and SSIM are out of range for all six groups of synthetic images. The average FID for the Perceptual Wasserstein loss is the closest to the original (0.82 vs. 0.80). Only the Wasserstein and Perceptual Wasserstein produce images that on average are within the range of the Blur metric for the original images.

Images generated with a Wasserstein and Perceptual Wasserstein loss are on average within the range of the original images for five out of seven metrics. Quantitatively, the Wasserstein and Perceptual Wasserstein networks generate images of a quality that most reassembles the original batch. In particular, the Perceptual Wasserstein loss generates the best results for four of the seven metrics. It appears to be the most appropriate loss optimization to generate cerebral organoid images with this AAE.

4.3. Psychovisual evaluation of synthetic images

In Figure 4, we compare the occurrence of each decision in percentages for *original* and *generated* groups. There is less misleading in original and generated group than right decisions. However, 30% of misleading is observed in the generated group. A misleading corresponds to a false positive answer.

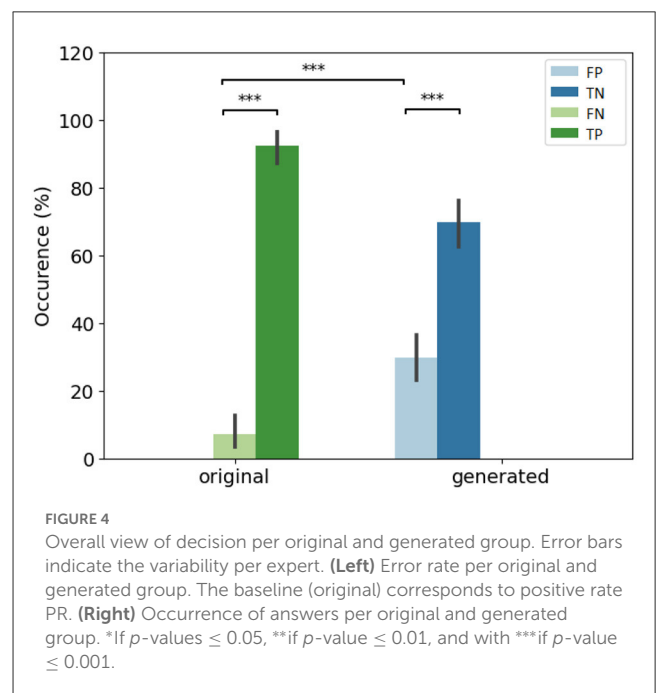
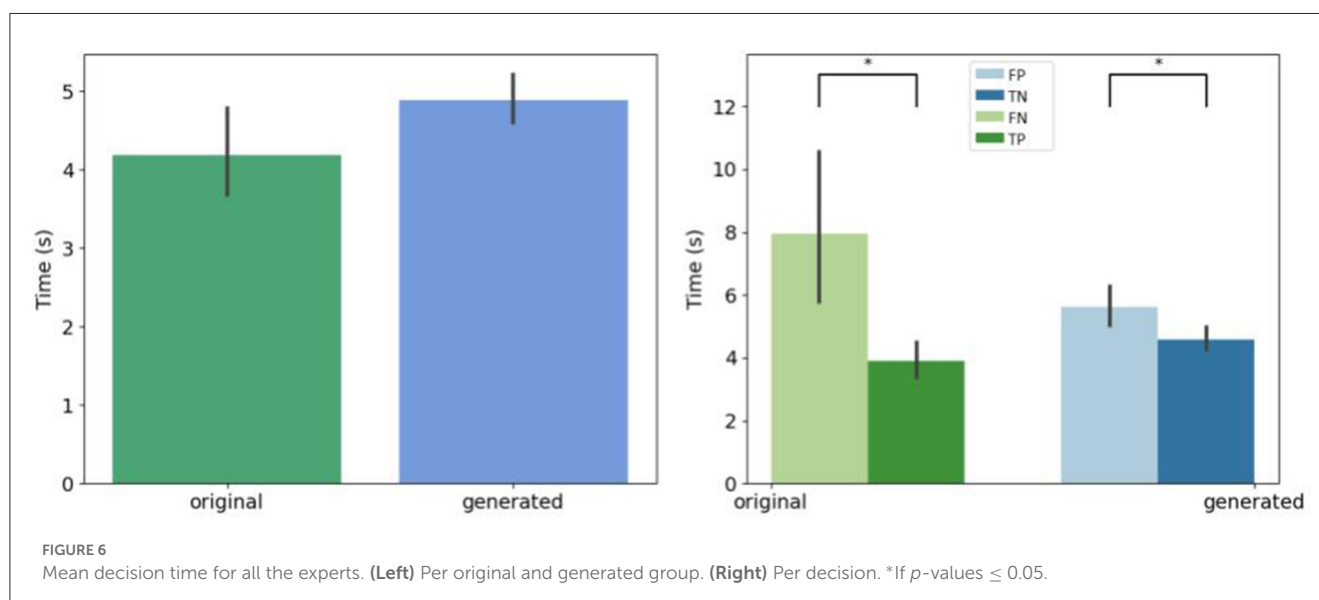
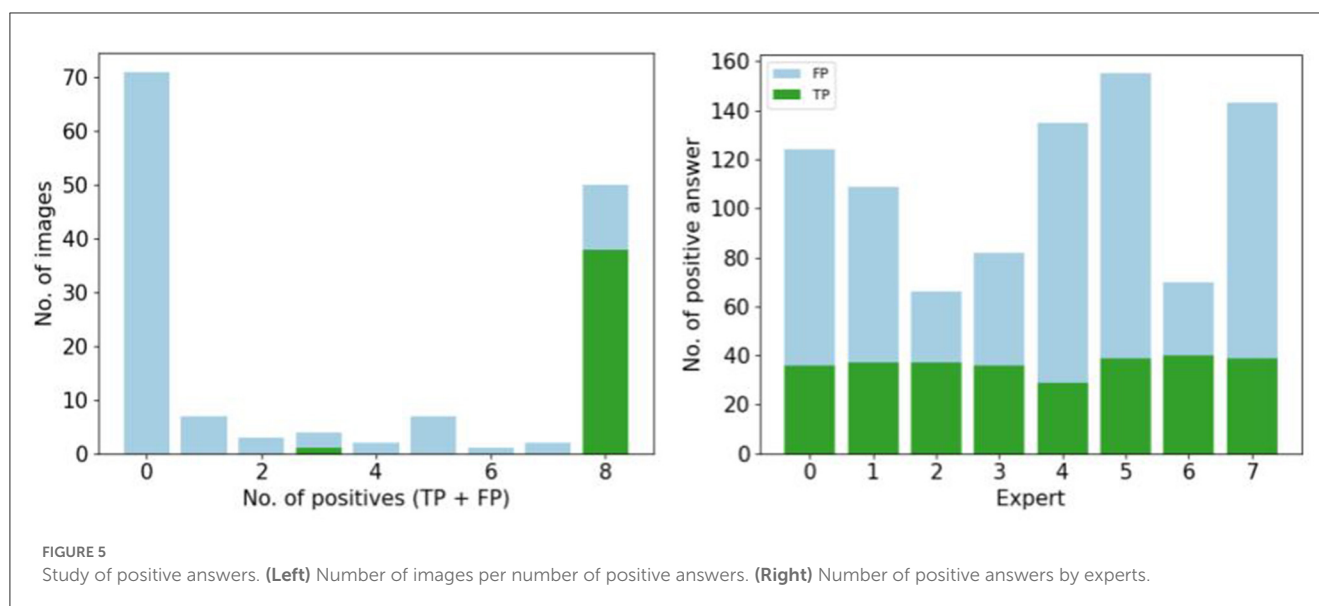


Figure 5 focuses on the images that are labeled as “natural” by the experts. We found the number of false positive selected images by all the participants is small (<20), and 30 images are selected by five participants. Almost 70 images are not selected at all by experts as natural (first column). To observe the number of false positive answer by each experts (see Figure 5). Three experts answer less positive answers than the others (less than the half of the visualized dataset). One expert considered over 150 images as natural.

We retrieve the decision time before the expert give an answer whatever the kind of generation, as shown Figure 6. Biological experts answer in the same time for generated and original images. However, the hesitation time is longer for false answers than for correct answers.

4.4. Feedback on the psychovisual procedure

Experts in cultures consider images as natural when an ovoid shape with neuroepithelial formation and some cell dispersion



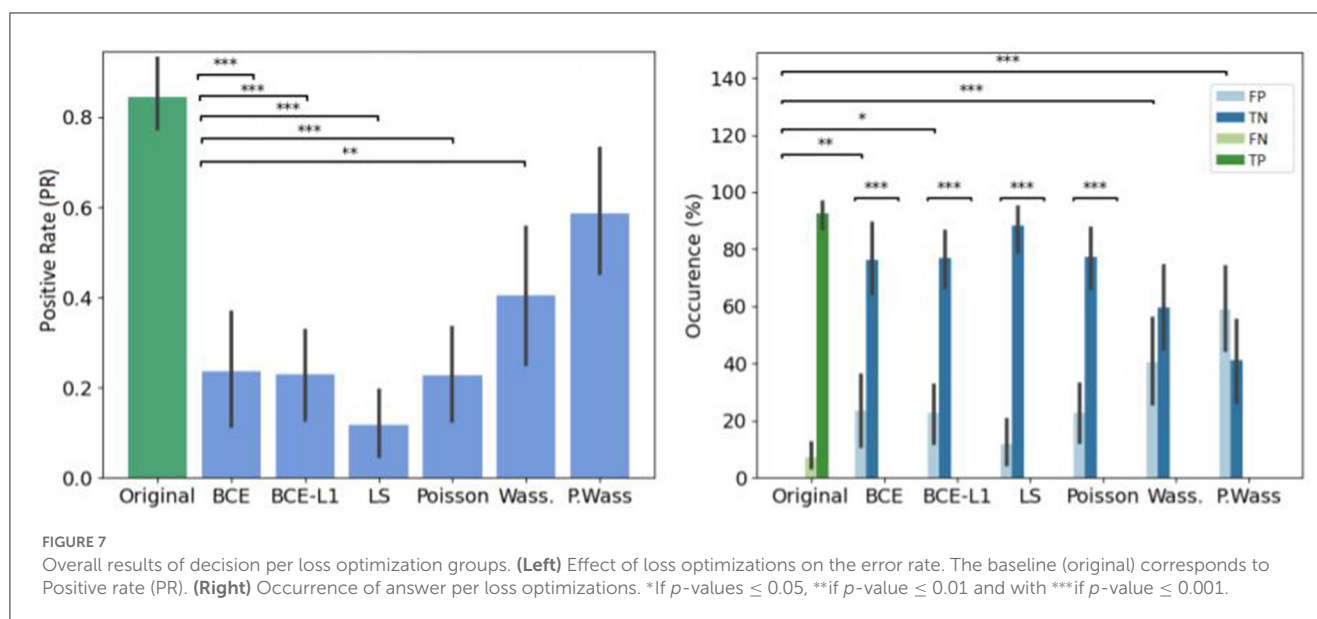
appears. When we ask experts why they classify a generated image as a synthetic, they answer the background contains an imprint, or superimposed contours or there is an artifact, or the image is too noisy, but they hesitate longer due to the possible content. They explain that they classify an original image as a synthetic because of a microscopic acquisition artifact, learned by one of the architectures, and reproduced on the worst synthetic images. They would like to have less images in a session, and a larger image on the screen.

4.5. Psychovisual evaluation of loss optimization

The error rate is particularly higher in the Wass. and P. Wass groups than for the original one (see Figure 7 left). In the P. Wass case this high score is strengthened by the absence of statistical

differences. These particular loss optimizations drive the experts to mislead and consider the images from these two groups as natural. In Figure 7 right, there is a difference between false positives of original and generated images from BCE, BCE+L1, Wass. and P. Wass. loss optimization. However, if we consider the intra-factor loss comparison, we can observe statistical differences between FP and TN of each for the BCE, BCE + L1, LS and Poisson loss rendering too small the proportion of misleading. There is no differences between these two occurrences decision for images generated by a Wass. or a P. Wass. loss showing 42% of FP and almost 60% of FP.

To observe which group of images is the most selected as positive, we observe the number of image selected by group in Figure 8 left, P. Wass. and Wass. images are selected as natural by the most of experts (a few Poisson, and a few BCE by seven experts and L1 + BCE). The same three experts as in Figure 5 only answer positive in most of the case for P. Wass. images (see Figure 8



right). Four experts answer more synthetic images but more even for P. Wass. Only one expert seems to answer identically for all synthetic group of images.

If we do not consider the kind of decision, there is no difference of decision time per group of synthesis, (Figure 9 left). When we study the decision time per group, the experts take more time to answer only when they are confronted to synthetic images generated with a least square optimization (Figure 9 right).

4.6. Concordance of metrics and psychovisual evaluations

4.6.1. Comparisons

After observing the psychovisual decisions by generated groups, we compare qualitative and similitude metrics to the previous results in order to verify if the same groups are selected, but also to verify if some metrics or combination of metrics can be used as a proxy to human psychovisual evaluation.

An overview of these results is given in Table 3 shows no differences between decision whatever the group of loss optimization or the calculated metric. FID is the highest for *Poisson* loss than for others groups whatever the kind of decision. BLUR metric is the highest for the decision with LS generated images. In term of SSIM, UQI indexes and PSNR, the decision reach the highest score for original images and no improvement is visible with generative methods. For MSE and MI the decision rate reach the highest score similarly for original and P. Wass. generated images. No differences are visible in term of decision with UQI.

4.6.2. Correlations

To identify the metrics which best correspond to the psychovisual evaluation, we plot metrics against error rate in Figure 10. We show the Blur scatterplot as an example of point representations. In this graphic, we observe that the green color

points (*Original*) are near the darker-purple ones (P. Wass.). Figure 10 also represents the KL divergence between the P. Wass group and all other groups for all metrics. If we look at the first column, P. Wass and *Original* images are closest according to SSIM, MI and UQI. These metrics are good candidates to build a metric that mimics psychovisual evaluations.

Figure 11 summarizes the correlations between the psychovisual assertion and the hesitation time or error rate according to each metric or chosen metric combinations. The main result is the absence of correlation for single metrics, however, combination with a Blur metric, FID (for the error rate correlations) and SSIM-FID-MI (for the time correlation) render the highest results in Figure 11 top left. The same result is represented in Figure 11 top right. To observe the group representation between the Error rate or the Time and the KL divergence of points represented for these two combinations (Blur-FID and Blur-SSIM-FID-MI) (see Figure 11 bottom right and left). The LS and BCE group are far from the others point representations. P. Wass. group is superimposing the original one with Wass. Others groups are not distinguishable, however, there are at the peripheral zone of the perceptual-original amount. The KL divergence representation with respect to the error rate of these two metric combinations is given in the bottom part of Figure 11.

4.7. Influence of psycho-validated images on a segmentation task

Then the second task is to verify the interest of using synthetic images which have been validated by 0, 2, 4, 6, or 8 biological experts to train a segmentation task.

4.7.1. Qualitative results

To observe the quality of the segmentation, we show a ground truth segmentation performed with the ITK-SNAP software, and

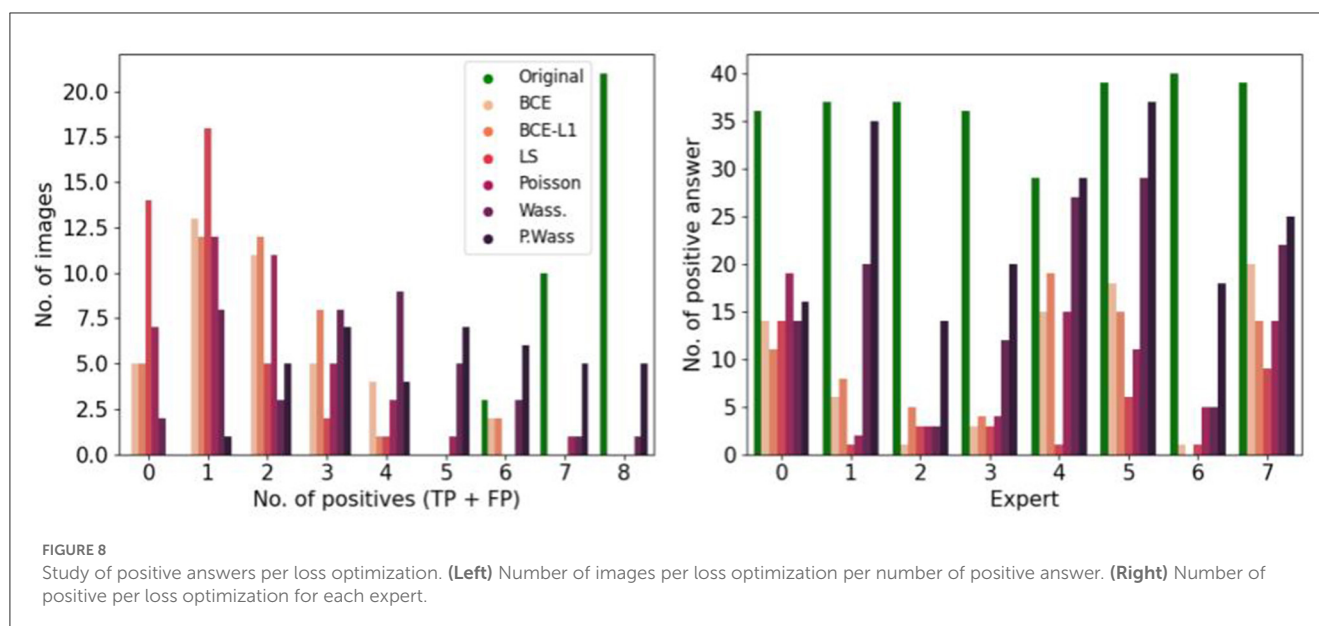


FIGURE 8

Study of positive answers per loss optimization. **(Left)** Number of images per loss optimization per number of positive answer. **(Right)** Number of positive per loss optimization for each expert.

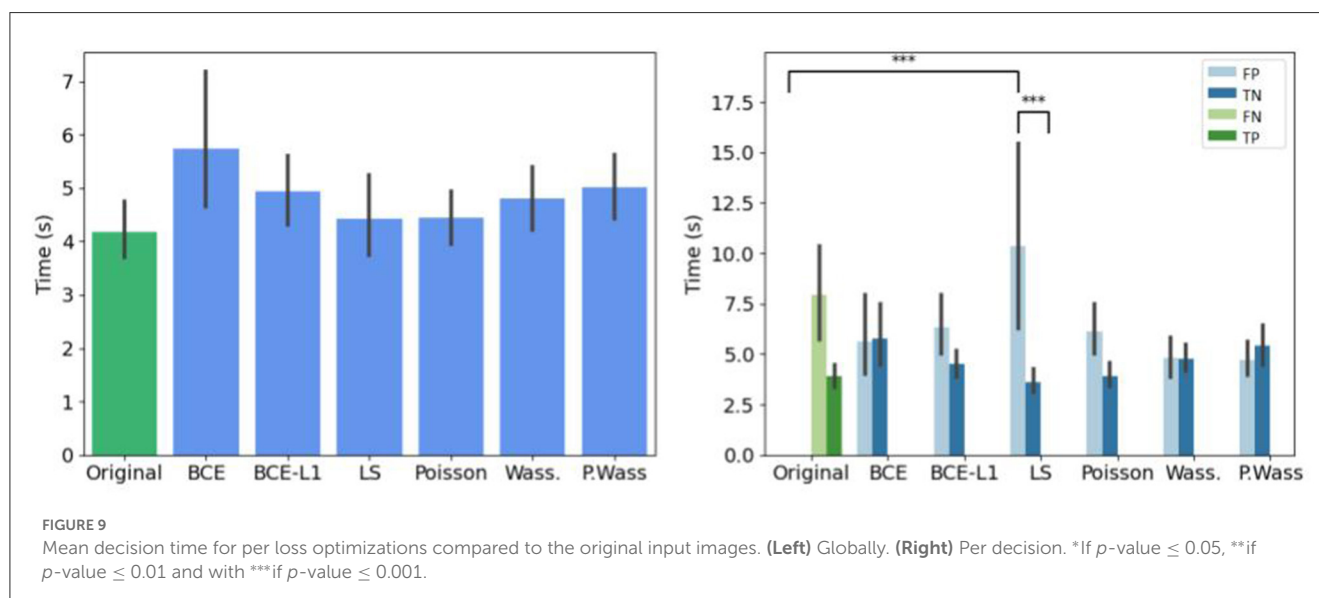


FIGURE 9

Mean decision time for per loss optimizations compared to the original input images. **(Left)** Globally. **(Right)** Per decision. *If p -value ≤ 0.05 , **if p -value ≤ 0.01 and with ***if p -value ≤ 0.001 .

automatic segmentations performed with a U-Net architecture with various data augmentation strategies, see vignettes in Table 4. The “0 experts” group corresponds to training the segmentation with images that have been selected by none of the experts. In the others training images are previously selected by 2, 4, 6, and 8 experts. We observe less false positive regions (in pink) and false negative regions (in green) if the segmentation is performed after training on a dataset containing images validated by 6 or 8 experts. If synthetic images selected by six or more experts are used for training, we observe almost no errors on the segmentation.

4.7.2. Quantitative results

Table 4 summarizes the metrics used to compare the ground truth segmentation with the segmentation performed by U-net trained on data augmented with classic strategies or on a varying portion of synthetic images. The segmentation is better when the

network is trained with images validated by 8 experts, with higher levels of Dice, Accuracy, Sensibility and F1-score. The highest sensitivity is reached by the group trained on 41 synthetic images validated by two experts and by the group trained on images validated by no experts. However, the corresponding specificity is very low.

5. Discussion

In this part we present to our knowledge the first psychovisual and metric evaluation comparison of Loss optimized generative adversarial network of brain organoid bright-field images. This study helps at validating most natural images generated by various AAE loss optimizations. We also contribute to strengthen metric evaluation by highlighting some images from optimized generated adversarial network to be perceived by Human biological expert as natural microscopic images: with a P. Wass. loss perception

TABLE 3 Average and standard deviation of five of metrics on original and synthetic images per psychovisual decision (true or false) and per loss optimization.

	Decision	Count	Blur		SSIM		PNSR		MSE		MI		UQI	
			avg.	σ	avg.	σ	avg.	σ	avg.	σ	avg.	σ	avg.	σ
Positive														
Original	TP	297	64.15	50.28	0.78	0.17	42	41.49	2973	3163	0.90	0.46	0.86	0.12
BCE	FP	79	100.22	15.40	0.62	0.06	13.52	3.20	3713	2610	0.856	0.14	0.83	0.10
BCEL1	FP	77	104.16	38.89	0.61	0.07	13.54	2.94	3549	2320	0.82	0.13	0.83	0.09
LS	FP	39	155	44.78	0.59	0.07	13.77	3.00	3403	2276	0.78	0.11	0.84	0.09
Poisson	FP	74	93.17	22.76	0.63	0.06	13.86	2.84	3275	2139	0.86	0.14	0.84	0.09
Wass.	FP	134	61.47	18.25	0.61	0.06	13.33	2.38	3499	1980	0.90	0.14	0.83	0.07
P. Wass.	FP	197	47.19	17.13	0.63	0.07	12.48	2.44	4272	2389	0.95	0.17	0.80	0.07
Negative														
Original	FN	23	44.93	25.36	0.71	0.01	10.10	0.90	6418	1326	0.88	0.23	0.78	0.04
BCE	TN	241	105.01	20.95	0.62	0.06	13.57	3.16	3647	2555	0.85	0.13	0.83	0.10
BCEL1	TN	243	110.43	39.30	0.60	0.07	13.60	2.87	3477	2248	0.81	0.13	0.83	0.09
LS	TN	281	156	44.21	0.58	0.06	13.79	2.80	3307	2118	0.78	0.11	0.84	0.09
Poisson	TN	246	106.12	25.34	0.62	0.06	13.86	2.77	3248	2090	0.84	0.13	0.84	0.09
Wass.	TN	186	56.96	17.18	0.63	0.07	13.41	2.58	3510	2116	0.90	0.13	0.83	0.07
P. Wass.	TN	123	41.22	14.04	0.63	0.07	12.69	2.55	4123	2414	0.95	0.16	0.80	0.075

Best values are displayed in bold.

neurobiologists are misled 60% of the time and 40% by Wass. They take more time to answer when they are misled. We compare human and metric evaluation and found mutual information to be the most related to their decision metric, although no correlation appeared in our experiment for single metrics, but only for combinations including blur. Using synthetic images validated by an increasing number of experts to train a segmentation network increases the accuracy of the segmentation with respect to classic augmentation strategies, even if the proportion of synthetic images decreases.

Synthetic images generated with AAE are coherent with original dataset and thus whatever the kind of loss optimization. The generation of brain organoid images with others architectures does not improve the synthesis in term of quality or similitude according to Brémond Martin et al. (2022a), whereas it seems the case with loss optimization. The P. Wass. loss optimization of AAE performs best according to metrics. Other loss optimizations show also high similitude, though with a lower quality. In this context, we plan to explore what type of information each loss brings during the image generation. We aim at trying others embedded losses (already used for segmentation tasks) during the generative process based upon high level prior like object shape, size topology or inter-regions constraints (El Jurdi et al., 2021). These losses could be used on condition that the morphological development of CO is better characterized.

Biomedical experts select around 40% of synthetic images as natural compared to the original dataset. Thus, the generation by AAE networks generate a large part of realistic images such as the background of bright-field acquisition or, their content. The

non-selected images where considered sometimes as non-natural due to some artifacts reproduced in some of them, or by a superposition of contours. Nevertheless, the selected images can help train a DL segmentation network.

A first argument of the strong validation of the selected images as natural is the time to take a decision (Shaffrey et al., 2002). If the time to answer natural for a generated group corresponds to the time to answer natural or original images, we could consider these two groups are perceived as similar. We found no differences between original and generated images and thus whatever the kind of loss optimization used to produce them. So they are not doubting when they classify an image as natural or generated. However psychovisual evaluation shows an increase of decision time before answering when they answer as false positive (depending of the loss optimization) or false negative. This behavior is specifically shown from a Least Square Loss Optimization generated images considered as natural. When we ask participants why they have doubt on a particular image, they answer that it was linked with some acquisition artifact learned by the generated process and found on a lot of images (a bunch of cells) or, by a blurry contour which could be due to the acquisition in the case of original images (Ali et al., 2022). For false negative answers, they only said that the artifact acquisition is also present (and they thought it was a generated). In the future, we think a pre-process image treatment has to be done on images to correct the acquisition artifact before the generative process, to avoid these false negative in the psychovisual evaluation or, to add a component in the generative network to avoid these artifacts (Galteri et al., 2017; Ali et al., 2022).

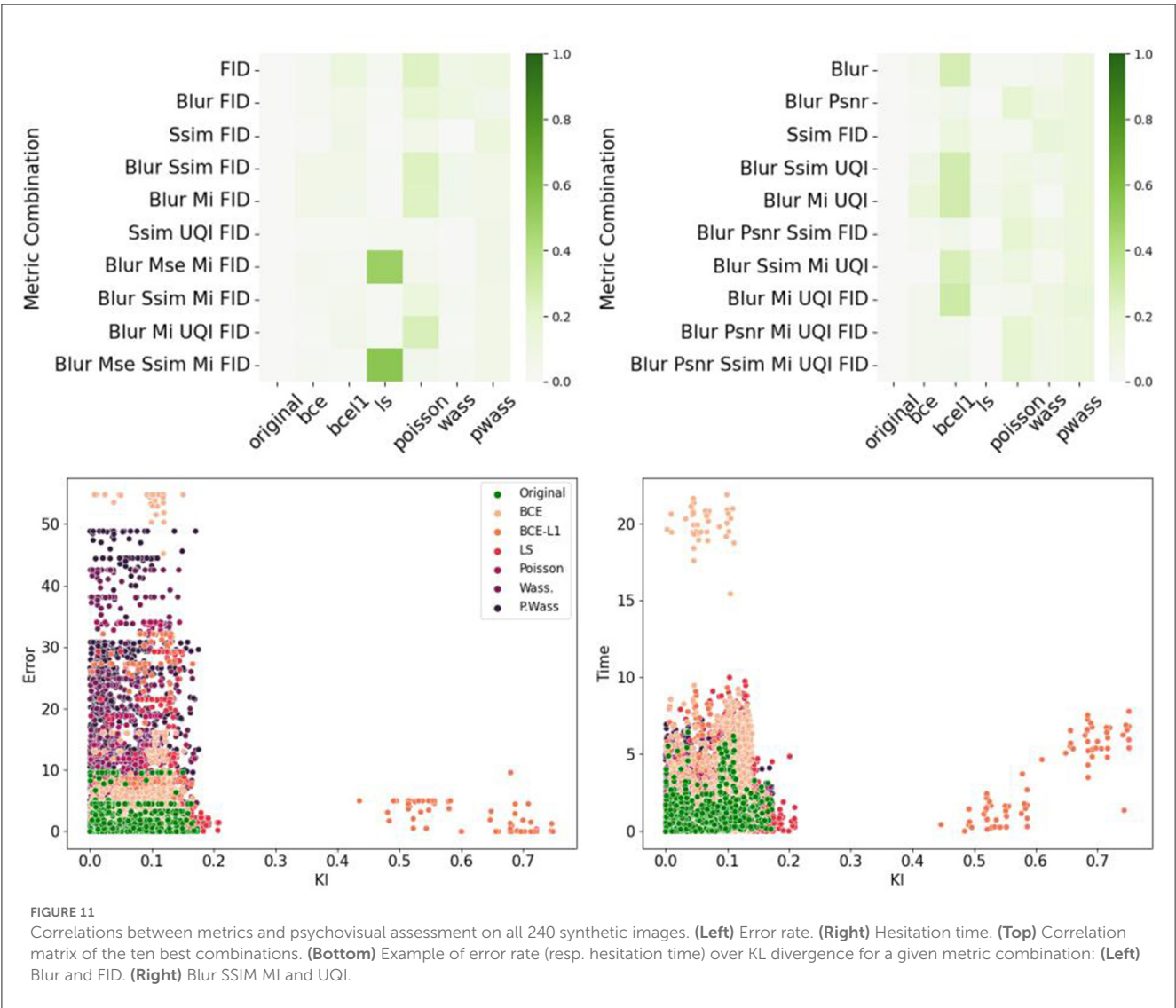
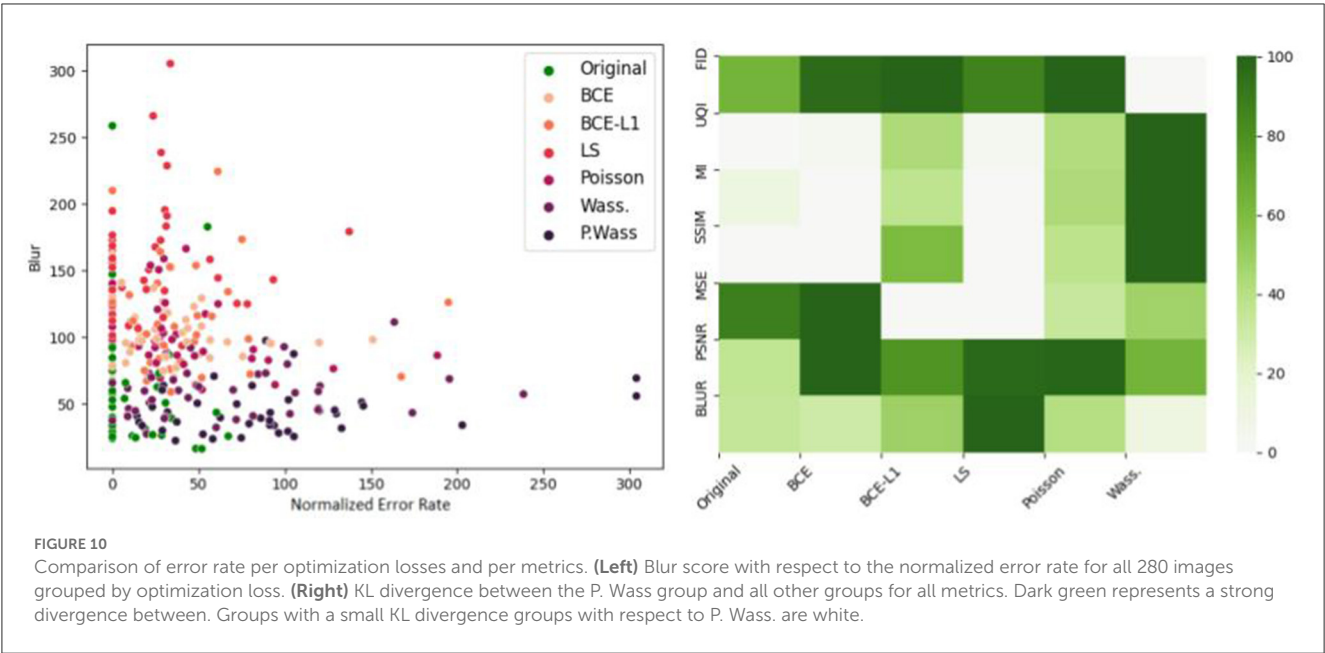


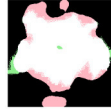
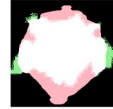





TABLE 4 Segmentation results after training on a varying number of synthetic images.

	GT	Classical	No. of experts validating the synthetic images				
			0	2	4	6	8
No. of synthetic images	–	0	33	41	22	16	14
Sample result							
Dice	1.0	0.80	0.62	0.59	0.61	0.67	0.82
Accuracy	1.0	0.77	0.63	0.61	0.68	0.62	0.93
Sensitivity	1.0	0.94	0.96	0.97	0.87	0.83	0.91
Specificity	1.0	0.92	0.50	0.48	0.60	0.56	0.94
F1-score	1.0	0.84	0.64	0.61	0.64	0.59	0.87

Seventy-nine images are used for training of which 39 are from the original data set. The remaining 40 are a combination of classic transformations and synthetic images. Tests performed with no synthetic images (classical) are compared to those performed with synthetic images validated with an increasing number of experts (0 to 8). Best values per metric are displayed in bold.

Only 15 images are selected by all the experts as natural. Five experts select over 100 images as natural and three <80. This study raises a question: could we use images considered by only five experts as natural in a training step? Thus we thought we need to increase the number of biological experts to overcome future studies in order to be more precise on the number of images considered as natural. We could also analyze the answers by the field of expertise of biological experts too (separate those who made only culture or only microscopic acquisition from those working in both fields).

Nonetheless, human Psychovisual experts choose in majority images from generative adversarial network as natural if they are from Wass. and P. Wass. loss and, a few BCE, BCE-L1. These two first kinds of generation are also highlighted by most of the metrics in an other study to have the better quality and to be the most similar to original images (Brémond Martin et al., 2022a). Thus, the psychovisual evaluation strengthen the choice of the use of these two and particularly the perceptual one in generative process. We can now confirm the idea that the regulation term of the Wass. distance between two images (Kupyn et al., 2018) could improve the learning of the pattern or characteristics of brain organoids in images and contribute to generate more natural images in term of content and aspects. In future studies we will remove the images that did not dupe the experts to only train on human-validated images. We would like to see if this increases the segmentation accuracy and study the impact on the morphological characterization.

However, the few BCE and BCE-L1 images selected as natural by psychovisual experts could maybe have also a great interest whereas the metric are not pointing them as natural images (Brémond Martin et al., 2022a). As we know the use of metric is still controversial for the GAN evaluation as they are measuring similitude and quality (Borji, 2019). Here, we could not highlight a strong correlation between the use of certain metrics and the decision to reject or not a generated image as natural. To correlate a metric and psychovisual evaluation instead of a binary answer “natural” or “not natural,” some authors use a graduation scale

(Pedersen and Hardeberg, 2012; Pedersen, 2015). This approach could be tested in future studies.

No metric used in this study could replace a human perceptual evaluation to decipher the naturality of an image generated. There is a certain link with FID, BLUR or MI and the group and MI with the mean decision but it remains weak. We could only say that similitude and referenced-bases metrics are more linked to the decision than qualitative metrics and non-reference-based metrics. And when we compare metrics with decisions some patterns appear according to the kind of loss optimization. The use or not of a measure to decipher natural generated examples is an issue recently discussed (Borji, 2019). To compare fairly images generated by various optimized models, there is no consensus for a use of a particular metric. In other fields such metric comparisons highlight a wavelet structural similitude index WSSI, a metric which based upon SSIM but less complex and more accurate in term of quality assessment (Rezazadeh and Coulombe, 2009; Pedersen, 2015). However, we do not want an identical image but one just resembling as a natural one. This could explain these metrics are not well designed for the GAN specific evaluation when they are considered alone. This study comparing the overall psycho-visual evaluation and seven metrics is one of the pioneer work which could contribute to help at pointing a metric of “natural,” and it failed partially.

Based upon our KL divergence maps, we suggest that a combination of metrics which best represents the psychovisual evaluation decision (BLUR, SSIM, MI, UQI) could be used as a substitue for a human psychovisual evaluation which is time consuming. Nevertheless, this work on metric combinations replacing a psychovisual evaluations need to be further studied. In other fields the combination of metrics help at pointing out some results in term of quality or similitude (Yao et al., 2005; Pedersen and Hardeberg, 2012; Okarma et al., 2021). An other idea could also to use non-reference quality metrics combinations (Rubel et al., 2022). Some authors tries also to implement directly a discriminator of generative adversarial networks based upon human perception, this could be a solution if it is not time consuming (Fujii et al., 2020; Arnout et al., 2021). It is not the case

in this study, for us, an important task is to find an appropriate metric for highlighting “the naturality” of the image and replacing the psycho-visual evaluation. An idea is to test psycho-metrics instead of classical similitude or quality metrics such as Hype from Zhou et al. (2019) which is an alternative of *FID* from Heusel et al. (2017), or implementing the GFI quality assessment created by Tian et al. (2022).

The more experts validate a portion of synthetic images in the dataset, the better the segmentation quality. This suggests that if more experts are available to select images, and strengthen the naturality of the synthetic dataset used during the training, it could improve the accuracy of the segmentation results. However, even if the psychovalidation of certain synthetic images allows us to improve the segmentation, this method is still subjective. It could include biases of the experts about the model, its configuration, and the project objective. It requires knowledge of which images are considered as natural and which ones are not for the target domain, so the number of experts available in the field diminished while our task required more experts. It is limited to the number of images that can be reviewed in a reasonable time (Booij et al., 2019).

Validation by six experts is the minimum to improve the segmentation, but quantitative analysis show us, eight experts validation is the minimum due to the equilibrium state between the specificity and the sensibility. The performance of human judges is not fixed and can improve over time, other articles choose a validation by 15 experts for instance which is not possible in our biomedical context which requires experts in the field (Denton et al., 2015; Salimans et al., 2016b).

We could also apply this psychovisual evaluation on others datasets to attempt to answer more specifically to the metric replacement. We thought about noise optimized generated images of brain organoids with an AAE for the same aim (Brémond Martin et al., 2022b). It could be interesting to observe if with a noise injection, similar to the bright-field acquisition images, generated images are more perceived as natural even if metrics are not pointing a particular kind of noise. Indeed, qualitative and similitude metrics point out Gaussian noise and shot noise injection. But as said previously, this could maybe only be due to the metric choice (Borji, 2019). An analysis of Psychovisual evaluation could maybe help at highlighting a combination of metrics. In this future study, it could be also interesting to observe for example the microscopic experience of the Biological expert as a new criterion. A larger application of this methodology could be made on others kind of generation (such as on GAN, Goodfellow et al., 2014 or DCGAN, Radford et al., 2015) and maybe help at pointing out the best GAN model for brain organoid generation used during a training segmentation task.

In this study we use a unbalanced dataset with more synthetic images than original. Nevertheless, biological experts do not know the number of real or synthetic images which render it unbiased. In future studies, we need to obtain and use more original images in order to re-equilibrate. We use a software created specifically for the psychovisual task for brain organoid images. The software needs to be updated due to some limitations. We have to realize batch process with pauses to limit the tiredness of biological experts similar to others psychovisual evaluations (Shaffrey et al., 2002). We

have to add also a cursor with a score instead of a button to estimate a natural range in future studies and facilitate correlations studies (Tian et al., 2022). The size of the image of the screen has to be increased but not for all the participants. Apart from these updates, the use of the software is simple and practical according to their feedback.

To strengthen our statistical analysis we should increase the number of biological experts. However, it could include biases of the experts: it requires knowledge of which image is considered as natural and which one is not for the target domain, so the number of experts available in the field is diminished while our task required more experts. The performance of human judges is not fixed and can improve over time, other articles choose a validation by 15 experts for instance which is not possible in our biomedical context which requires experts in the field (Denton et al., 2015; Salimans et al., 2016a). Moreover, psychovisual evaluation is limited to the number of images that can be reviewed in a reasonable time (Borji, 2019). The tradeoff between the number of synthetic images used to train a network and the number of validating experts could be further explored.

6. Conclusion

In this study psycho-visual evaluations allow us to:

- Validate some synthetic image generated from loss optimization of generative brain organoid images with an AAE in term of decision time and decision.
- Describe the quality and similitude of the synthetic images with the original dataset by a metric validation.
- Verify if some synthetic images could be considered as natural by psychovisual expert decision.
- Compare psychovisual and metric evaluations.
- Paves the way to finding a metric or a metric combination that mimics psychovisual evaluations.
- Show the interest of selecting images validated by the highest number of experts in a data augmentation strategy for a segmentation task.

This selected images could be used in the training phase of a segmentation task in order to help at their morphological development characterization for instance. We also need to evaluate psychovisually noise injected optimized synthesized images.

In future studies we suggest a combination of metrics or a perceptual metric could maybe help at replacing the psycho-visual assessment which is time consuming. Such methodology could be used for others brain organoid data-sets generated with a generative adversarial network.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

This article was an idea of CB-M, CS-C, CC, and AH. The literature search and data analysis was performed by CB-M. Graphics tables and figures were conceived by CB-M and CS-C. The first draft was written by CB-M while CS-C, CC, and AH critically revised the work. All authors contributed to the article and approved the submitted version.

Funding

ANRT supported CB-M Ph.D. scholarship (2019/0635).

Acknowledgments

The authors would like to thank Dr. Michel Boissière for his assistance and feedback. The authors would like to acknowledge all

the team from ERRMECe laboratory for participating to blinded psychovisual evaluations of synthetic images.

Conflict of interest

CB-M and CC were employed by Witsee, Neoxia.

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

Publisher's note

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

References

- Albanese, A., Swaney, J. M., Yun, D. H., Evans, N. B., Antonucci, J. M., Velasco, S., et al. (2020). Multiscale 3D phenotyping of human cerebral organoids. *Sci. Rep.* 10, 21487. doi: 10.1038/s41598-020-78130-7
- Ali, M. A. S., Hollo, K., Laasfeld, T., Torp, J., Tahk, M.-J., Rinken, A., et al. (2022). AttSeg—artifact segmentation and removal in brightfield cell microscopy images without manual pixel-level annotations. *Sci. Rep.* 12, 11404. doi: 10.1038/s41598-022-14703-y
- Alqahtani, H., Kavakli, M., and Kumar Ahuja, D. G. (2019). "An analysis of evaluation metrics of gans," in *International Conference on Information Technology and Applications*, Vol. 7 (Sydney, NSW).
- Arnout, H., Bronner, J., and Runkler, T. (2021). "Evaluation of generative adversarial networks for time series data," in *2021 International Joint Conference on Neural Networks* (Shenzhen: IEEE), 1–7. doi: 10.1109/IJCNN52387.2021.9534373
- Balaji, Y., Chellappa, R., and Feizi, S. (2019). "Normalized wasserstein for mixture distributions with applications in adversarial learning and domain adaptation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul: IEEE), 6499–6507. doi: 10.1109/ICCV.2019.00660
- Booij, T. H., Price, L. S., and Danen, E. H. J. (2019). 3D Cell-based assays for drug screens: challenges in imaging, image analysis, and high-content analysis. *SLAS Discov.* 24, 615–627. doi: 10.1177/2472555219830087
- Borji, A. (2019). Pros and cons of gan evaluation measures. *Comput. Vis. Image Underst.* 79, 41–65. doi: 10.1016/j.cviu.2018.10.009
- Brémond Martin, C., Simon Chane, C., Clouchoux, C., and Histace, A. (2021). Recent trends and perspectives in cerebral organoids imaging and analysis. *Front. Neurosci.* 15, 629067. doi: 10.3389/fnins.2021.629067
- Brémond Martin, C., Simon Chane, C., Clouchoux, C., and Histace, A. (2022a). "AAEGAN loss optimizations supporting data augmentation on cerebral organoid bright field images," in *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Volume 4, VISIGRAPP* (Setúbal: SciTePress), 307–314. doi: 10.5220/0010780000003124
- Brémond Martin, C., Simon Chane, C., Clouchoux, C., and Histace, A. (2022b). "AAEGAN optimization by purposeful noise injection for the generation of bright-field brain organoid images," in *Internal Conference On Image Processing Theory Tools and Application: Special Session Biological and Medical Image Analysis* (Salzburg: IEEE), 6. doi: 10.1109/IPTA54936.2022.9784149
- Chakradhar, S. (2016). New company aims to broaden researchers' access to organoids. *Nat. Med.* 22, 338–338. doi: 10.1038/nm0416-338
- Chen, Y., Yang, X.-H., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., et al. (2022). Generative Adversarial Networks in Medical Image augmentation: a review. *Comput. Biol. Med.* 144, 105382. doi: 10.1016/j.compbiomed.2022.105382
- Denton, E., Chintala, S., Szlam, A., and Fergus, R. (2015). "Deep generative image models using a laplacian pyramid of adversarial networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS'15* (Cambridge, MA: MIT Press), 1486–1494.
- El Jurdi, R., Petitjean, C., Honeine, P., Cheplygina, V., and Abdallah, F. (2021). High-level prior-based loss functions for medical image segmentation: a survey. *Comput. Vis. Image Underst.* 210, 103248. doi: 10.1016/j.cviu.2021.103248
- Fernandez, R., Rosado, P., Vegas, E., and Reverter, F. (2021). Medical image editing in the latent space of generative adversarial networks. *Intell.-Based Med.* 5, 100040. doi: 10.1016/j.ibmed.2021.100040
- Fujii, K., Saito, Y., Takamichi, S., Baba, Y., and Saruwatari, H. (2020). "HUMANGAN: generative adversarial network with human-based discriminator and its evaluation in speech perception modeling," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (Barcelona: IEEE), 6239–6243. doi: 10.1109/ICASSP40776.2020.9053844
- Galteri, L., Seidenari, L., Bertini, M., and Bimbo, A. D. (2017). "Deep generative adversarial compression artifact removal," in *IEEE International Conference on Computer Vision* (Venice: IEEE), 4836–4845. doi: 10.1109/ICCV.2017.517
- Gomez-Giro, G., Arias-Fuenzalida, J., Jarazo, J., Zeuschner, D., Ali, M., Possemis, N., et al. (2019). Synapse alterations precede neuronal damage and storage pathology in a human cerebral organoid model of CLN3-juvenile neuronal ceroid lipofuscinosis. *Acta Neuropathol. Commun.* 7, 222. doi: 10.1186/s40478-019-0871-7
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. *Adv. Neural Inf. Process. Syst.* 3.
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. (2017). "Improved training of wasserstein gans," in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (Red Hook, NY), 5769–5779.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (Red Hook, NY: Curran Associates Inc), 6629–6640.
- Hu, B., Tang, Y., Chang, E. I.-C., Fan, Y., Lai, M., Xu, Y., et al. (2019). Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks. *IEEE J. Biomed. Health Inf.* 23, 1316–1328. doi: 10.1109/JBHI.2018.2852639
- Jiang, M., Zhi, M., Wei, L., Yang, X., Zhang, J., Li, Y., et al. (2021). Fa-gan: fused attentive generative adversarial networks for mri image super-resolution. *Comput. Med. Imaging Graph.* 92, 101969. doi: 10.1016/j.compmedimag.2021.101969
- Joyce, J. M. (2011). "Kullback-Leibler divergence," in *International Encyclopedia of Statistical Science*, ed M. Lovric (Berlin: Springer Berlin Heidelberg), 720–722. doi: 10.1007/978-3-642-04898-2_327

- Kelava, I., and Lancaster, M. A. (2016). Dishing out mini-brains: current progress and future prospects in brain organoid research. *Dev. Biol.* 420, 199–209. doi: 10.1016/j.ydbio.2016.06.037
- Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., and Matas, J. (2018). “DeblurGAN: blind motion deblurring using conditional adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 8183–8192. doi: 10.1109/CVPR.2018.00854
- Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., et al. (2020). Generative adversarial networks and its applications in biomedical informatics. *Front. Public Health* 8, 164. doi: 10.3389/fpubh.2020.00164
- Lancaster, M. A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L. S., Hurler, M. E., et al. (2013). Cerebral organoids model human brain development and microcephaly. *Nature* 501, 373–379. doi: 10.1038/nature12517
- Mahmood, F., Borders, D., Chen, R. J., McKay, G. N., Salimian, K. J., Baras, A., et al. (2018). Adversarial training for multi-organ nuclei segmentation in computational pathology images. *IEEE Trans. Med. Imaging* 39, 3257–3267. doi: 10.1109/TMI.2019.2927182
- Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B. (2016). Adversarial autoencoders. *arXiv*. [preprint]. doi: 10.48550/arXiv.1511.05644
- Malm, P., Brun, A., and Bengtsson, E. (2015). Simulation of brightfield microscopy images depicting papsmear specimen. *Cytometry Part A* 87, 212–226. doi: 10.1002/cyto.a.22624
- Mao, X., Li, Q., Xie, H., Lau, R. Y. K., Wang, Z., Smolley, S. P., et al. (2017). “Least squares generative adversarial networks,” in *IEEE International Conference on Computer Vision* (Venice: IEEE), 2813–2821. doi: 10.1109/ICCV.2017.304
- Okarma, K., Lech, P., and Lukin, V. V. (2021). Combined full-reference image quality metrics for objective assessment of multiply distorted images. *Electronics* 10, 2256. doi: 10.3390/electronics10182256
- Pedersen, M. (2015). “Evaluation of 60 full-reference image quality metrics on the CID: IQ,” in *IEEE International Conference on Image Processing* (Quebec City, QC: IEEE), 5. doi: 10.1109/ICIP.2015.7351068
- Pedersen, M., and Hardeberg, J. (2012). Full-reference image quality metrics: classification and evaluation. *Found. Trends Comput. Graph. Vis.* 7, 1–80. doi: 10.1561/06000000037
- Plum, J., Maintz, J., and Viergever, M. (2003). Mutual-information-based registration of medical images: a survey. *IEEE Trans. Med. Imaging* 22, 986–1004. doi: 10.1109/TMI.2003.815867
- Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv*. [preprint]. doi: 10.48550/arXiv.1511.06434
- Rezagadeh, S., and Coulombe, S. (2009). “A novel approach for computing and pooling structural similarity index in the discrete wavelet domain,” in *2009 16th IEEE International Conference on Image Processing (ICIP)* (Cairo: IEEE), 2209–2212. doi: 10.1109/ICIP.2009.5413886
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *arXiv*. [preprint]. doi: 10.48550/arXiv.1505.04597
- Rubel, A., Ieremeiev, O., Lukin, V., Fastowicz, J., and Okarma, K. (2022). Combined no-reference image quality metrics for visual quality assessment optimized for remote sensing images. *Appl. Sci.* 12, 1986. doi: 10.3390/app12041986
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., et al. (2016a). “Improved techniques for training gans,” in *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS’16* (Red Hook, NY: Curran Associates Inc), 2234–2242.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., et al. (2016b). Improved techniques for training gans. *Adv. Neural Inf. Process. Syst.* 29.
- Shaffrey, C. W., Jermyn, I. H., and Kingsbury, N. G. (2002). Psychovisual evaluation of image segmentation algorithms. *Adv. Concepts Intell. Vis. Syst.* 7.
- Singh, N. K., and Raza, K. (2021). “Medical image generation using generative adversarial networks,” in *Health Informatics: A Computational Perspective in Healthcare*, eds R. Patgiri, Biswas A., and P. Roy (Singapore: Springer Nature Singapore), 19. doi: 10.1007/978-981-15-9735-0_5
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., et al. (2016). Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans. Med. Imaging* 35, 1299–1312. doi: 10.1109/TMI.2016.2535302
- Tian, Y., Ni, Z., Chen, B., Wang, S., Wang, H., Kwong, S., et al. (2022). Generalized visual quality assessment of gan-generated face images. *arXiv*. [preprint]. doi: 10.48550/arXiv.2201.11975
- Tsomo, E., Kim, H. J., Paik, J., and Yeo, I.-K. (2008). Efficient method of detecting blurry images. *Journal of Ubiquitous Convergence Technology* 2, 27.
- Wang, Z., and Bovik, A. (2002). A universal image quality index. *IEEE Signal Process. Lett.* 9, 81–84. doi: 10.1109/97.995823
- Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861
- Wargnier-Dauchelle, V., Simon-Chane, C., and Histace, A. (2019). Retinal blood vessels segmentation: improving state-of-the-art deep methods. computer analysis of images and patterns. CAIP 2019. *Commun. Comput. Inf. Sci.* 1089, 5–16. doi: 10.1007/978-3-030-29930-9_1
- Yao, S., Lin, W., Ong, E., and Lu, Z. (2005). “Contrast signal-to-noise ratio for image quality assessment,” in *IEEE International Conference on Image Processing 2005* (Genova: IEEE), I-397. doi: 10.1109/ICIP.2005.1529771
- Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: a review. *Med. Image Anal.* 58, 101552. doi: 10.1016/j.media.2019.101552
- Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C., et al. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31, 1116–1128. doi: 10.1016/j.neuroimage.2006.01.015
- Zhao, H., Gallo, O., Frosio, I., and Kautz, J. (2017). Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* 3, 47–57. doi: 10.1109/TCI.2016.2644865
- Zhou, S., Gordon, M. L., Krishna, R., Narcomey, A., Morina, D., Bernstein, M. S., et al. (2019). Hype: human eye perceptual evaluation of generative models. *LCTR*, 15.

Frontiers in Neuroscience

Provides a holistic understanding of brain
function from genes to behavior

Part of the most cited neuroscience journal series
which explores the brain - from the new eras
of causation and anatomical neurosciences to
neuroeconomics and neuroenergetics.

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

