IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 May 2024, accepted 7 June 2024, date of publication 11 June 2024, date of current version 15 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3412961

RESEARCH ARTICLE

Hybrid Graph Representation Learning for Carotid Artery Stenosis Detection Based on Multimodal Retinal OCTA Images

WENTING LAN[1], JINKUI HAO[2], SHENGJUN ZHOU[3], JINGFENG ZHANG[4], SHAODONG MA[2], AND YITIAN ZHAO[2], (Member, IEEE)

[1]Department of Radiology, First Affiliate Hospital of Ningbo University, Ningbo 315211, China
[2]Laboratory of Advanced Theranostic Materials and Technology, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315000, China
[3]Department of Neurosurgery, First Affiliate Hospital of Ningbo University, Ningbo 315211, China
[4]Department of Radiology, Ningbo No. 2 Hospital, Ningbo 315099, China

Corresponding authors: Shengjun Zhou (nbzhoushengjun@126.com), Jingfeng Zhang (jingfengzhang73@163.com), and Shaodong Ma (mashaodong@nimte.ac.cn)

This work was supported in part by the National Science Foundation Program of China under Grant 62272444; in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR22F020008LTGY23H180003; in part by the Key Research and Development Program of Zhejiang Province under Grant 2023C04017; in part by Ningbo 2025 S&T Megaprojects under Grant 2022Z134, Grant 2022Z127, and Grant 2021Z134; and in part by the Key Project of Ningbo Public Welfare Science and Technology under Grant 2023S012 and Grant 2021S107.

ABSTRACT Carotid artery stenosis (CAS) is one of the major causes of cerebral ischemic stroke. Rapid and precise detection of CAS is crucial for early intervention and reducing ischemic stroke incidence. Neuroimaging techniques, as the gold standard for evaluating cerebral abnormalities in CAS, suffer from limitations including expensive and time-consuming, hindering their use in large-scale screening. The ophthalmic artery is a branch of the internal carotid artery, several studies suggest that the biomarkers on retinal optical coherence tomography angiopraphy (OCTA) images are associated with CAS. Thus, retinal OCTA as a non-invasive and high-resolution imaging technique has potential as a suitable approach for identifying CAS patients. In this work, we developed a hybrid graph-based deep learning model to detect CAS from OCTA images. Given the differential impact of CAS on arteries and veins, we explicitly leverage the artery and vein information within the retinal region to enhance the sensitivity of the model to the change in microvasculature. We construct a hybrid graph representation by combining arterial and venous features, with the aim of improving the model's ability to extract and integrate diverse anatomical information for more accurate CAS detection. For evaluation, we enrolled 182 CAS and 239 control subjects in this study. The experimental results demonstrated our retinal image analysis-based AI model, received promising results in distinguishing CAS and control subjects, with AUC of 0.7765 and an accuracy of 0.7750.

INDEX TERMS Carotid artery stenosis, deep learning, retinal image, GNN, multi-modal.

I. INTRODUCTION

Carotid artery stenosis (CAS), a leading cause of ischemic stroke and eye diseases, poses a major global public health challenge [1]. Delayed diagnosis and treatment exacerbate patients' cognitive decline and substantially raise the risk of falling. Thus, early CAS detection is critical for preventing severe complications [2]. While positron emission tomography (PET) perfusion imaging represents the gold standard for assessing cerebral perfusion [3], its high costs and radiation risks preclude large-scale screening and rapid CAS detection.

As the retina shares embryological origins and microvascular properties with the brain [4], retinal abnormalities reflect cerebral microvascular diseases [5]. Optical coherence tomography angiography (OCTA) offers high-resolution retinal vascular imaging, enabling visualization of $5\mu m$

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

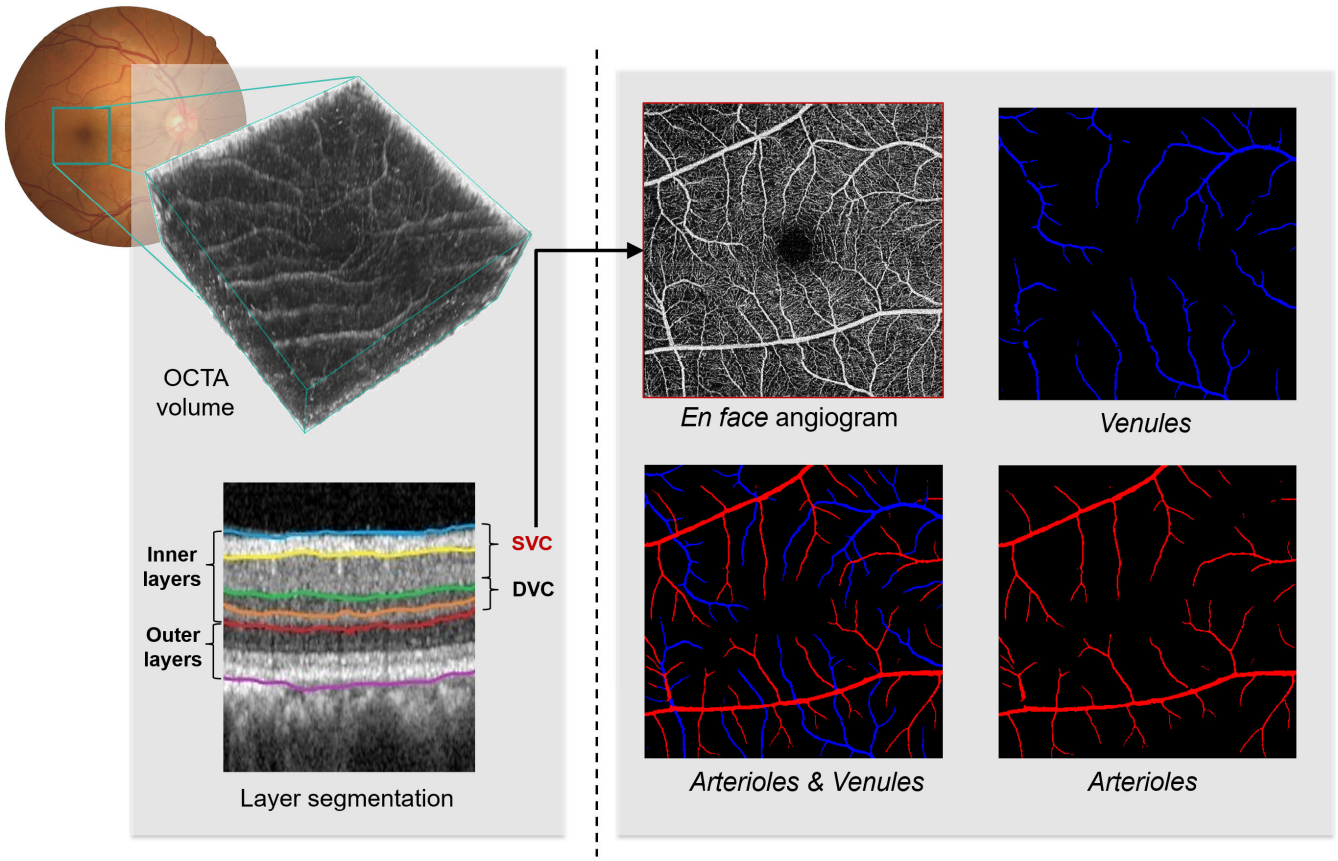


FIGURE 1. Example of OCTA volume and *en face* angiogram of superficial vascular complex (SVC). We used existing methods to extract arterioles and venules of the retina, to improve the sensitivity of the model for the microvasculature change.

capillaries for potential CAS biomarkers, as shown in Fig. 1. Several studies [6], [7], [8], [9] have reported a relationship between biomarkers obtained from retina and CAS. Therefore, developing an OCTA-based automated algorithm is a potential tool for rapid CAS detection.

However, unlike eye diseases such as diabetic retinopathy and age-related macular degeneration, which have visible lesions on the retina, the impact of CAS on retinal vascular is hard to observe. For example, existing findings show that the vessel density in CAS patients is significantly lower than controls [6], [10], and vessel tortuosity is also significantly decreased in CAS cases [8]. Consider emphasizing the above features are hard to observe, the key to build an effective deep learning model is how to capture the fine-grained change of microvasculartur globally and locally. Meanwhile, the ophthalmic artery, which supplies blood to the retina, is a branch of the internal carotid artery. Thus, structural changes in the internal carotid artery may have different effects on arterioles and venules of the retina [10]. Capturing these intricate relationships requires modeling beyond individual modalities. It is important to fuse complementary information from arterioles and venules through cross-modal connections, thereby improving feature learning and classification performance.

In this study, we propose a hybrid graph representation framework for CAS detection using OCTA images. By explicitly leveraging the information on retinal arterioles and venules, the model can focus on the subtle changes of the microvasculature, thus improving the detection performance. In addition, we construct a hybrid graph combining the spatial and semantic graph to model the features of microvasculartur from different levels. Moreover, we employ a graph attention network with position embedding to extract and fusion features from a hybrid graph and complete the final prediction. The experimental results have shown that the proposed model can provide trustworthy and interpretable results, and thus provides a potential tool for the rapid screening of CAS and the analysis of associations between retinal structures and CAS diseases.

II. RELATED WORK

A. DISEASE DETECTION ON OCTA

OCTA has been widely used in computer-aided eye-related disease detection and analysis due to ites ability to provide 3D depth-resolved high-resolution angiographic images of the retina [11]. It is mainly used to diagnose different types of eye disease including glaucoma [12], age-related macular degeneration (AMD) [13], [14], and mostly diabetic retinopathy

(DR) [15], as well as neurodegenerative diseases [16], [17], [18]. These existing automatic diagnostic methods can be divided into two categories, conventional methods based on machine learning and deep-learning-based methods. For example, to classify no DR and DR, Liu et al. [19] applied four machine learning methods with OCTA texture features, and experiments showed that logistic regression regularized with the elastic net penalty performs the best compared with the SVM, XGBoost, and logistic regression itself. Alfahaid and Morris [20] proposed a local texture features-based algorithm for AMD classification over OCTA images. As an extension, they utilized a rotation-invariant uniform local binary pattern (LBP) descriptor and Principal Component Analysis (PCA) to capture local decor-related texture patterns [21].

Different from conventional machine learning-based methods, deep learning methods can obtain more informative features from OCTA images, which enable more accurate diagnoses of diseases. For example, [22] established an AI model using ResNet34 to classify different types of AMD (normal, dry AMD, and wet AMD with either active or inactive CNV) using choriocapillaris *en face*. In order to utilize the depth information of OCTA, some studies proposed using the *en face* combination of different retinal slabs as input. Reference [23] applied an ensemble learning method to classify DR using multiple *en face* images of OCTA. Zang et al. [24] proposed a CNN architecture for the DR classification using different *en face* images. Reference [25] developed a CNN-based algorithm to screen DR using OCTA images. They compared the performance of the individual *en face* image and the combined data by concatenation, and found that the combined data performed worse than the deep capillary plexus *en face* alone. Although the complementary information provided by different projection maps has a potential benefit for classification tasks, how to mine and exploit their relationships requires further and in-depth exploration. Existing methods usually utilize the multiple *en face* of OCTA through early or middle-stage information fusion based on CNNs. Even though CNN can efficiently extract local features through parameter-sharing convolution kernels, it is difficult to learn and fully utilize the relationship between different instances.

B. GRAPH REPRESENTATION LEARNING FOR DISEASE DETECTION

Compared with CNN, GNN can take advantage of the relationship between nodes to capture the common patterns/biomarkers. On this basis, there has been an increasing focus on GNN in computer vision [26], [27], [28], [29], [30] and automatic disease diagnosis [31], [32], [33], [34], [35]. Some studies are based on non-image data [36], [37], [38], [39], [40]. For example, [37] introduced a GNN model for disease prediction based on electronic medical records (EMR), in which external knowledge bases are utilized to augment the insufficient EMR data. Hao et al. [41] proposed an uncertainty-guided graph attention network (UG-GAT) for disease diagnosis in CT images. Each subject is represented as a graph, where nodes represent the image features of slices, and edges encode the spatial relationship between them. Chen et al. [42] proposed a GNN framework to analyze CT images, similar to [41], where they treated each slice as a graph node and predicted the categories of each graph at patient level.

In addition, methods [43], [44], [45] often construct a large graph using all the population data, in which each node represents the image features of an individual subject. Reference [45] designed a graph framework that leveraged both imaging and non-imaging information for AD analysis. Its nodes are associated with image features of subjects, while phenotypic information is integrated as edge weights. Jiang et al. [44] proposed a hierarchical GNN model for brain network learning with network topology and subject's association simultaneously. In sharp contrast with previous works, our work constructs a multilevel graph representation at instance-level as well as subject-level to capture both the intra-instance and the inter-instance relationship.

III. MATERIALS AND METHODS

A. MATERIALS

Our study enrolled unilateral asymptomatic or symptomatic CAS patients from the local hospitals. This study was approved by the ethics committee of the Institute of Biomedical Engineering, Chinese Academy of Sciences, and adhered to the principles of the Declaration of Helsinki.

The inclusion criterion for participants was carotid artery stenosis of at least 50% confirmed on CT angiography (CTA). Exclusion criteria were: 1) Ocular surgery within the past 6 months; 2) Myopia over 6 diopters; 3) Neurodegenerative diseases including Alzheimer's, Parkinson's, and multiple sclerosis; 4) Cerebral hemorrhage on MRI; 5) Normal cerebral perfusion in the ipsilesional middle cerebral artery territory, defined as $\leq$ 30% relative difference in cerebral blood flow between bilateral MCA territories on CTP maps; 6) Non-atherosclerotic intracranial stenosis; 7) Previous intervention or surgery in the intra- or extra-carotid artery; 8) Retinal or choroidal diseases such as age-related macular degeneration, cataracts, or glaucoma.

This dataset comprised 182 OCTA subjects with CAS and 239 control subjects. There was no significant difference in age between the CAS and the control groups ($P > 0.1$). All participants underwent imaging using a swept-source OCT angiography system (VG200; SVision Imaging) with a 200,000 A-scan/second scan rate and 1050nm central wavelength. The scan area was 6×6mm centered on the fovea, generating en face angiograms of the superficial vascular complex through automatic segmentation to assess retinal microvascular perfusion. Retinal arterioles and venules were then extracted using our previous topology-aware method [46].

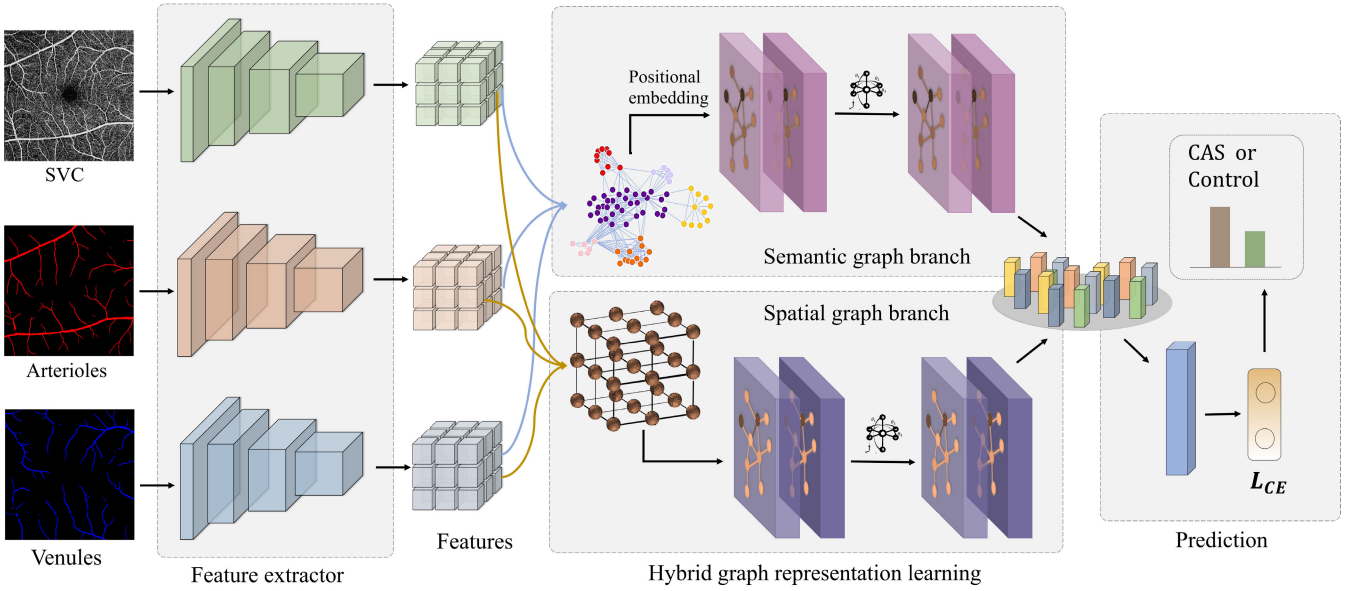


FIGURE 2. Detailed design of our model. The model extracts features from different regions using a CNN and represents each region as a node in the graph. Two types of graphs are constructed: a spatial graph and a semantic graph. Graph attention networks (GATs) are used to extract features from the graphs, and a multilayer perceptron fuses the hybrid graph information.

B. PROPOSED METHOD

Given an input OCTA image I_o and corresponding arteriole map I_a and venule map I_v, as shown in Fig. 2, we aim to predict whether the subject corresponding to the input is with CAS. We proposed a hybrid graph representation learning model,that first extracted the features of different regions using a CNN and considering each region as a node in the graph. We considered two different relationships to construct different graph representations: a spatial graph based on the locations of different regions and instances, and a semantic graph based on the similarity of features. We then utilized two separated graph attention networks (GATs) to extract features from the two graph representations. Finally, a multilayer perceptron network was used to fuse hybrid graph information.

We adopted ConvNext [47] as the backbone for our deep-learning model. ConvNext is a four-stage convolutional neural network with two convolutional layers and one max pooling layer in each stage. It uses batch normalization and ReLU activation to improve training stability and performance. We used a multi-branch model to extract features from the original image, arteriole map and venule map. Each branch of our model consists of a ConvNext backbone, and we then constructed a hybrid graph based on these features.

1) HYBRID GRAPH REPRESENTATION

The input image $X^k \in X$ of size $h \times w$ is partitioned into $n \times n$ regions. Each region is regarded as a node of the graph, in which node features are obtained by CNN. We describe $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ as a graph representation with $|\mathbb{V}| = n^2$ nodes and $|\mathbb{E}|$ edges. $|\cdot|$ represents the cardinality of a given set. For each $v_i \in \mathbb{V}$, h_i is the corresponding F-dimensional feature vector. Let $H \in \mathbb{R}^{n^2 \times F}$ be the node feature matrix, and $A \in \mathbb{R}^{n^2 \times n^2}$ be the sparse adjacency matrix encoding the edge connections between the nodes.

We constructed two types of graph representations: a spatial graph $\mathbb{G}_{sp}$ and a semantic graph $\mathbb{G}_{se}$. $\mathbb{G}_{sp}$ and $\mathbb{G}_{se}$ have same initial node features $H \in \mathbb{R}^{n^2 \times F}$, and different adjacency matrix $A_{sp} \in \mathbb{R}^{n^2 \times n^2}$ and $A_{se} \in \mathbb{R}^{n^2 \times n^2}$. The sensitivity of the retinal microvasculature to CAS may vary from region to region, and consideration of the relationship between different regions is essential for a comprehensive assessment of retina changes. In the spatial graph, we first built the connections in a single input according to the location of different regions in Euclidean space. Adjacent region in the original image will have an edge in the spatial graph. Then, we considered the relationship of multiple inputs: an edge is added between the corresponding regions of the original image and arteriole/venule images. In addition to spatial relationships, we also try to capture semantic correlations between different features, i.e., representations of disease-related features, so that similar features can mutually reinforce each other. Henceforth, we proceed to construct the semantic graph by evaluating the similarity between each node and other nodes, subsequently establishing a weighted graph based on these similarity metrics.

2) GRAPH REPRESENTATION LEARNING AND PREDICTION

To mine the relationship of different nodes from two types of graphs, we use the GATs to update the node features. Then we can fuse the information of different graph representations. Since the operation of graph convolution

TABLE 1. Classification results of different methods. The best performance is highlighted in boldface.

Method	Accuracy	AUC	Precision	F1-score	Kappa
Early fusion	0.7000	0.7297	0.7266	0.6802	0.3684
Middle fusion	0.7090	0.7506	0.7234	0.6883	0.3632
Late fusion	0.7250	0.7411	0.7474	0.7106	0.4240
MCC	0.7000	0.7424	0.7546	0.6700	0.3617
MUCO	0.7250	0.7343	0.7520	0.7135	0.4387
GCN	0.7250	0.7443	0.7520	0.7135	0.4387
GAT	0.7555	0.7692	0.7926	0.7356	0.4648
UG-GAT	0.7555	0.7236	0.7926	0.7356	0.4648
Ours	**0.7750**	**0.7765**	**0.8073**	**0.7632**	**0.5287**

is concerned with local connectivity, the correspondence between semantic and spatial graphs is not taken into account. To better perform the next step of feature fusion, we add the position embedding to enhance GATs with structural information. By adding the same position embedding to the spatial and semantic graphs, the correspondence between the two can be preserved during the graph convolution process, which facilitates the subsequent information fusion. Specifically, given a graph $\mathbb{G}^k = (\mathbb{V}, \mathbb{E})$ with a set of node features $h = \left\{\overrightarrow{h}_1, \overrightarrow{h}_2, \ldots, \overrightarrow{h}_{n^2}\right\}, \overrightarrow{h_i} \in \mathbb{R}^F$, positional embedding first transforms a vector of node features into a new vector of node features. Then the GAT layer updates the node features and obtains the new embeddings, i.e., $h' = \left\{\overrightarrow{h}'_1, \overrightarrow{h}'_2, \ldots, \overrightarrow{h}'_{n^2}\right\}, \overrightarrow{h}'_i \in \mathbb{R}^{F'}$, with F' being the dimension of the updated node feature. The details for the update are as follows. Firstly, a linear transformation parameterized by a shared weight matrix $W \in \mathbb{R}^{F' \times F}$ is employed for each node. Then we calculate the attention coefficients e_{ij} using the self-attention operation:

$$e_{ij} = LeakyReLU(\overrightarrow{a}^T[W\overrightarrow{h}_i + p_i \parallel W\overrightarrow{h}_j + p_j]), \quad (1)$$

where p_i is the positional embedding for node i, $e_{ij} \in E$ indicates the attention value of node j to node i, and $E \in \mathbb{R}^{n^2 \times n^2}$ is the attention coefficient matrix. $\parallel$ is the concatenation operation, $\overrightarrow{a}^T \in \mathbb{R}^{2F'}$ is a learnable weight vector implemented by a fully-connected layer followed by the LeakyReLU activation (with negative input slope $\alpha = 0.2$).

Here we utilize a learnable positional embedding, and spatial graph $\mathbb{G}_{sp}$ and semantic graph $\mathbb{G}_{se}$ share the same embedding to ensure the alignment of the node information during the feature fusion. Then we normalize the coefficient e'_{ij} using the softmax function to make it comparable across different nodes:

$$\alpha_{ij} = Softmax_j(e_{ij}) = \frac{exp(e_{ij})}{\sum_{l \in N_i} exp(e_{il})}, \quad (2)$$

where N_i denotes the neighborhood of node i in the graph, showing that only e'_{ij} for neighboring node $j \in N_i$ is considered during updating the node feature to avoid involving irrelevant nodes. Finally, we use the normalized attention coefficients α_{ij} to calculate a weighted sum of the involved node features to obtain the updated features for each node:

$$\overrightarrow{h}'_i = ELU(\sum_{j \in N_i} \alpha_{ij} W \overrightarrow{h}_j), \quad (3)$$

where ELU represents the exponential linear unit (ELU) nonlinearity. After a two branches account for spatial graph $\mathbb{G}_{sp}$ and semantic graph $\mathbb{G}_{se}$, we obtain the hybrid representation by cascading the output features of two branches. Finally, we use a fully-connected layer to produce the final prediction.

C. IMPLEMENTATION DETAILS

All experiments were implemented using PyTorch and run on a workstation with 1 NVIDIA GeForce 3090 GPU. We utilized ConvNext as the backbone of the embedding extractor in CNN branch. The two branches of spatial and semantic graphs are both based on the GAT layer. The GAT layers have two attention heads computing 1024-dimensional features. The Adam optimizer is used to optimize the model with a weight decay of 0.0005 and a batch size of 4. The initial learning rates were set to 0.0001, which gradually decayed to zero after 400 epochs using a Cosine annealing scheduler. In addition, data augmentation including random horizontal flip and vertical flip was employed to enlarge the training dataset.

During the model evaluation, we employed five-fold cross-validation for more reliable analysis over these relatively small datasets. We utilized commonly used metrics for multi-class classification to assess performance. These metrics include Accuracy, Area under the ROC Curve (AUC), balanced precision, balanced F1-score and kappa analysis.

IV. RESULTS

A. CLASSIFICATION PERFORMANCES

We selected several methods for comparison that can be used for multi-modal inputs, including several CNN-based methods: early fusion [48], middle fusion [49], late

TABLE 2. Performance comparison of different feature fusion methods. Our proposed attention mechanism fusion approach outperforms simple feature addition, feature concatenation, and conventional attention mechanisms, achieving superior results across multiple evaluation metrics.

Method	Accuracy	AUC	Precision	F1-score	Kappa
Addition	0.6333	0.6800	0.7639	0.6370	0.3265
Concatenation	0.7090	0.6509	0.6911	0.6704	0.2294
Self-attention	0.7666	0.7320	0.7805	0.7459	0.4444
Ours	0.7750	0.7765	0.8073	0.7632	0.5287

fusion [23], MCC [50], CRBM [51], and MUCO [52], as well as several GNN-based methods: GCN [53], GAT [54], and UG-GAT [41]. In early fusion, the features from different modalities are fused together before being input into the classification model. This fusion is typically achieved by concatenating or stacking the features of different modalities. In middle fusion, the representations learned from each modality are then fused at an intermediate stage, typically between layers or modality-specific layers of the models. In late fusion, the features from different modalities are separately input into their respective models for processing and learning. The classification results from each modality are obtained independently, and the final decision is made by fusing these results. MCC achieves multimodal fusion by incorporating a correlation constraint and a dual attention-based fusion block into the model. CRBM computes multiple representations by considering the consistency among representations from different modal. By stacking correlation RBMs, a correlation DBN is created to learn unified representations that fuse multimodal together. MUCO can simultaneously consider the correlation and complementarity between different modalities, thereby improving classification accuracy. The GCN-based method treats images from different modalities as separate graph nodes, transforming image classification into a graph classification problem. The GAT-based method utilizes multi-head self-attention, taking into account the differences in contributions from different nodes. UG-GAT proposes that different modalities have varying contributions to classification and utilizes uncertainty as guidance, considering the contribution of different modalities during the classification process.

The quantitative results are presented in Table 1. First of all, our model outperformed all the compared CNN-based/GNN-based methods in all five evaluation metrics, including an accuracy of 0.7750, a balanced precision of 0,8073, an F1-score of 0.7632, as well as kappa and AUC values of 0.5287 and 0.7765, respectively. Secondly, the specifically designed *en face* fusion approach, known as MUCO, outperformed general fusion approaches, indicating that leveraging the relationships between different *en face* images is crucial for obtaining accurate classification results. Thirdly, GNN-based methods demonstrated superior performance compared to CNN-based methods. This is likely due to the fact that GNNs are more effective in modeling and extracting relationships between nodes, and leveraging the correlation and complementarity between different *en face* images is essential for accurate classification.

B. EFFECTIVENESS OF HYBRID GRAPH FUSION

In this section, we investigate the significance of hybrid graph representation learning in terms of result effectiveness. Specifically, we compare simple feature addition, feature concatenation, and attention mechanism fusion with our method. It is evident that the attention mechanism outperforms simple feature addition, concatenation, and the aforementioned methods. As shown in Table 2, our proposed method surpasses these approaches, achieving superior performance across multiple evaluation metrics. The attention mechanism fusion approach that we proposed can be seen as an extension of conventional attention mechanisms. However, in contrast to common attention mechanisms, our method incorporates prior knowledge of spatial and semantic relationships. By leveraging this prior knowledge, we enable a more effective fusion of features, leading to enhanced performance. The integration of spatial and semantic relationships as prior knowledge allows our method to capture and exploit the inherent dependencies and correlations among graph elements. This enriched understanding of the graph structure enables more informative and discriminative feature fusion. As a result, our method surpasses the standalone attention mechanism fusion and achieves remarkable improvements in various evaluation metrics.

C. EFFECTIVENESS OF MULTIPLE INPUTS

We stated the importance of exploiting the information of arterioles and venules images, which may conduce to a more accurate result. In this section, we compare the performances when using the SVC only and using multiple inputs. Table 3 shows the performance of our model with different inputs. It can be observed that multi-input outperforms SVC only in all metrics, indicating that multiple inputs can provide more information to the model, leading to more accurate and reliable results. By considering both arterioles and venules images and harnessing the power of deep learning, we can better capture the intricate changes in retinal vasculature associated with CAS. This approach holds promise for improving diagnostic accuracy and advancing our understanding of the condition.

TABLE 3. Ablation study and effectiveness of multiple inputs.

Method	Accuracy	AUC	Precision	F1-score	Kappa
w/o spatial or semantic	0.7250	0.7343	0.7520	0.7135	0.4387
w/o semantic or semantic	0.7200	0.7532	0.7498	0.7001	0.4027
w/o positional embedding	0.7556	0.7692	0.7926	0.7356	0.4648
SVC map only	0.7000	0.7462	0.7546	0.67008	0.3617
Arteriole map only	0.6909	0.6625	0.7154	0.6515	0.2857
Venule map only	0.7272	0.6570	0.7729	0.6925	0.3617
Ours	0.7750	0.7765	0.8073	0.7632	0.5287

D. ABLATION STUDY

Ablation studies were conducted to validate the effectiveness of different components of our proposed hybrid graph representation learning model. We performed several experiments by systematically removing or modifying components and observing the impact on model performance, as shown in Table 3. We compared the performance of our model with and without the hybrid graph representation. Removing either the spatial or semantic graph resulted in a significant drop in performance, demonstrating the importance of integrating both types of graph information. Furthermore, we evaluated the role of positional embeddings by training our model without them. The results showed a decrease in accuracy and precision, indicating that positional embeddings play a crucial role in aligning spatial and semantic information and enhancing feature fusion.

E. VISUALIZATION ANALYSIS

To understand the decision-making mechanism of our model, we conducted an interpretability analysis using visualization techniques. The results are presented in Fig. 3 It is evident that the visualization results of CAS samples and control samples exhibit distinct patterns. CAS samples exhibit higher activations and larger regions of significance compared to control samples. Conversely, control samples show lower activation levels. Additionally, differences in activations and visualized regions can be observed among the SVC, arterioles, and venules images, indicating that different inputs may provide complementary information for the detection results.

The variations observed in the visualized activations and regions of significance offer valuable insights into the decision-making process of our model. The higher activations and larger significant regions in CAS samples suggest that the model focuses on specific areas that are indicative of the presence of CSA. These visualized patterns align with our understanding of CSA-related changes in retinal vasculature, highlighting the model's ability to capture disease-specific features.

Furthermore, the differences observed in the activations and visualized regions among the various inputs: SVC, arterioles, and venules images, indicate that each input modality contributes unique and complementary information to the detection process. The arterial and venous images capture different aspects of the retinal vasculature, enabling a more comprehensive assessment of its structural and functional characteristics. By considering multiple inputs, our model can leverage the complementary information from these different modalities, leading to improved detection performance.

V. DISCUSSION

OCTA is a non-invasive imaging technique that can be performed quickly and without the need for contrast agents. Our automated analysis framework enhances the efficiency of CAS screening by providing rapid and accurate assessments, potentially reducing the reliance on more invasive and costly procedures like positron emission tomography (PET) perfusion imaging. By leveraging high-resolution OCTA images and advanced machine learning algorithms, our approach enables the early detection of CAS, which is critical for preventing severe complications such as ischemic strokes and cognitive decline. Early intervention can significantly improve patient outcomes and reduce healthcare costs.

However, integrating AI-based diagnostic tools into clinical settings requires robust IT infrastructure and seamless integration with existing electronic health records (EHR) systems. Ensuring compatibility and data interoperability is essential for the successful adoption of these technologies. Extensive training and validation are required to ensure the reliability and generalizability of AI models. Large and diverse datasets from different populations and clinical settings are needed to validate the performance of our proposed method.

Our proposed method builds upon and extends existing research in several ways. Firstly, compared to traditional methods that often focus on single-modality data or simple fusion techniques, our approach employs a hybrid graph model that integrates spatial and semantic information from multiple vascular structures, providing a more comprehensive analysis. Furthermore, we utilize state-of-the-art graph attention networks with positional embeddings, which enhance the model's ability to capture and fuse relevant features. This leads to significant improvements in classification

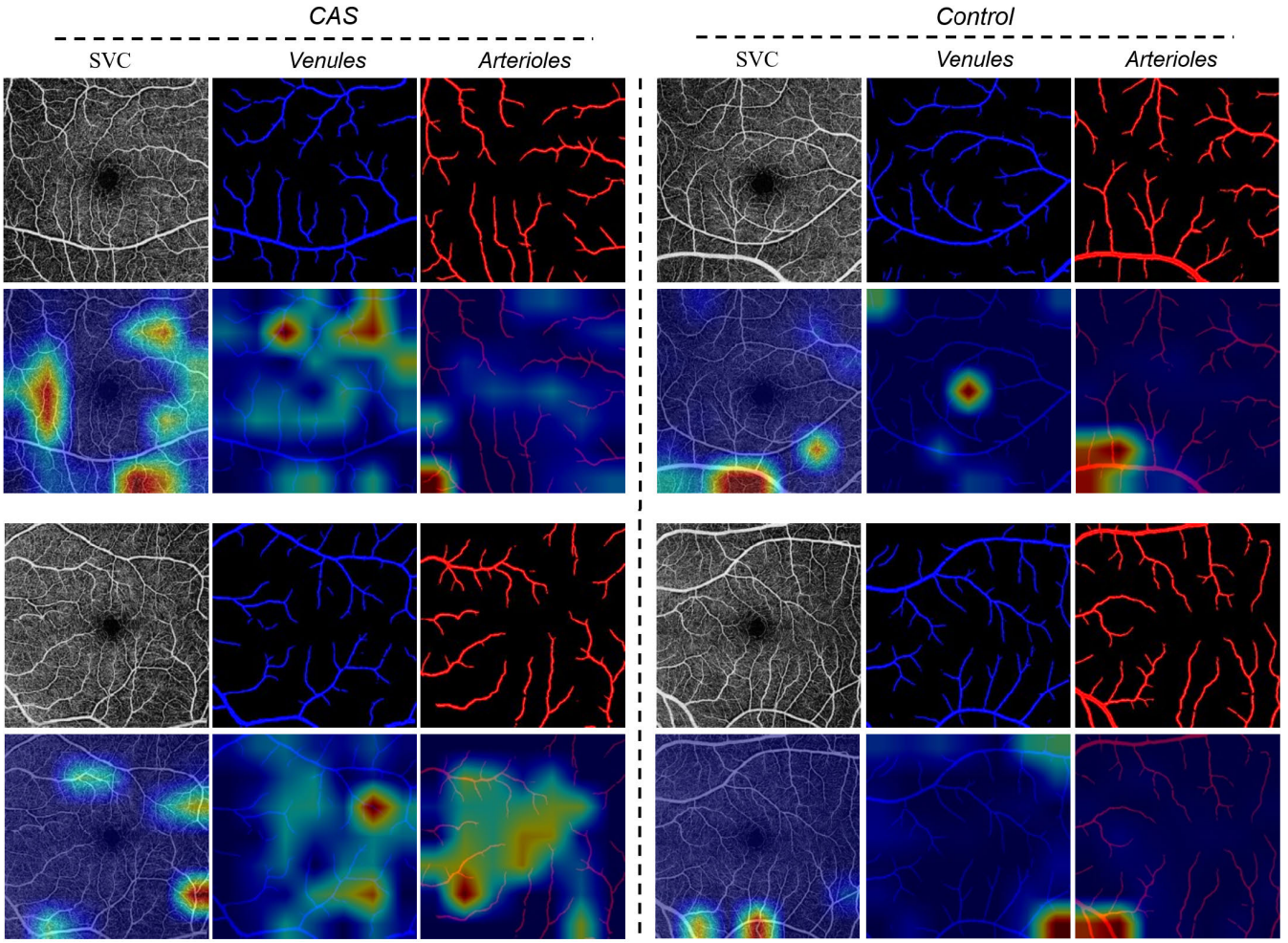


FIGURE 3. Visualization results showing distinct patterns between CAS and control samples. CAS samples exhibit higher activations and larger regions of significance, while control samples display lower activation levels. Variations in activations and visualized regions among SVC, arterioles, and venules images suggest complementary information for detection.

performance compared to other graph-based methods like GCN and traditional CNN-based fusion approaches.

We chose graph-based methods over other neural networks (NNs) and traditional methods due to graph-based methods excel at capturing the relationships between different regions in the OCTA images. Retinal microvasculature exhibits intricate spatial and semantic relationships. Graph-based methods, such as Graph Attention Networks (GATs), are inherently designed to capture these complex inter-node relationships, which is challenging for traditional convolutional neural networks (CNNs) and other NNs. The spatial graph representation helps in preserving the positional context, while the semantic graph representation captures feature similarity, both of which are critical in identifying subtle vascular changes associated with CAS. In addition, the use of attention mechanisms in GATs allows the model to focus on the most relevant nodes and edges, which improves the detection performance by highlighting the critical regions of interest in the retinal images. This is particularly beneficial for CAS detection, where the changes might be subtle and localized. Moreover, our approach involves combining information from arterioles and venules, which requires an effective fusion of multi-modal data. Graph-based methods excel in integrating heterogeneous data sources, making them ideal for leveraging complementary information from different vascular structures. We have included these explanations in the revised manuscripts.

To improve the generalizability and robustness of our model, we plan to expand our dataset by including more diverse populations from multiple clinical centers. This will help validate the model across different clinical setting and make it more applicable to real-world clinical settings. While our current focus is on CAS, the methodologies developed in this study can be extended to detect other vascular diseases that affect the retinal microvasculature. Future research will explore applications in conditions such as diabetic retinopathy, hypertensive retinopathy, and other neurovascular disorders. Additionaly, We can explore the use

of other modalities, such as structural OCT, to further enhance the accuracy and robustness of our model.

VI. CONCLUSION

In this work, we have proposed a novel hybrid graph representation learning framework for automated carotid artery stenosis detection using multimodal OCTA imaging. The key innovation lies in constructing spatial and semantic graphs to model the retinal microvasculature from global and local perspectives. Graph attention networks are then utilized to integrate the multilevel graph features. Our experiments demonstrated superior classification performance compared to CNN and GNN baselines, with the proposed method achieving significantly higher AUC. Visualization analysis also provided valuable insights, showing the model focuses on disease-relevant regions and highlighting the complementary information provided by the SVC, arteriole, and venule images. Overall, this work presents a promising approach for exploiting multimodal retinal imaging data to improve the detection of carotid artery stenosis. The graph-based representation learning paradigm allows effective feature extraction and integration across different imaging modalities and levels. Future work involves expanding the dataset size, investigating additional graph construction techniques, and incorporating other retinal imaging modalities like OCT. From a clinical perspective, an important next step is translating the automated detection system into a usable screening tool that can aid in the early diagnosis and prevention of severe complications from carotid artery stenosis.

ACKNOWLEDGMENT

(Wenting Lan and Jinkui Hao contributed equally to this work.)

REFERENCES

[1] T. Blaser, K. Hofmann, T. Buerger, O. Effenberger, C.-W. Wallesch, and M. Goertler, "Risk of stroke, transient ischemic attack, and vessel occlusion before endarterectomy in patients with symptomatic severe carotid stenosis," *Stroke*, vol. 33, no. 4, pp. 1057–1062, Apr. 2002.

[2] J. Schröder, M. Heinze, M. Günther, B. Cheng, A. Nickel, T. Schröder, F. Fischer, S. S. Kessner, T. Magnus, J. Fiehler, A. Larena-Avellaneda, C. Gerloff, and G. Thomalla, "Dynamics of brain perfusion and cognitive performance in revascularization of carotid artery stenosis," *NeuroImage, Clin.*, vol. 22, Jun. 2019, Art. no. 101779.

[3] M. V. Mattoli, G. Treglia, M. L. Calcagni, A. Mangiola, C. Anile, and G. Trevisi, "Usefulness of brain positron emission tomography with different tracers in the evaluation of patients with idiopathic normal pressure hydrocephalous," *Int. J. Mol. Sci.*, vol. 21, no. 18, p. 6523, Sep. 2020.

[4] L. Erskine and E. Herrera, "Connecting the retina to the brain," *ASN Neuro*, vol. 6, no. 6, Oct. 2014, Art. no. 175909141456210.

[5] S. J. Wiseman et al., "Retinal capillary microvessel morphology changes are associated with vascular damage and dysfunction in cerebral small vessel disease," *J. Cerebral Blood Flow Metabolism*, vol. 43, no. 2, pp. 231–240, Feb. 2023.

[6] X. Li, S. Zhu, S. Zhou, Y. Zhang, Y. Ding, B. Zheng, P. Wu, Y. Shi, H. Zhang, and H. Shi, "Optical coherence tomography angiography as a noninvasive assessment of cerebral microcirculatory disorders caused by carotid artery stenosis," *Disease Markers*, vol. 2021, pp. 1–10, Jul. 2021.

[7] X. Liu, B. Yang, Y. Tian, S. Ma, and J. Zhong, "Quantitative assessment of retinal vessel density and thickness changes in internal carotid artery stenosis patients using optical coherence tomography angiography," *Photodiagnosis Photodynamic Therapy*, vol. 39, Sep. 2022, Art. no. 103006.

[8] L. Pierro, A. Arrigo, M. De Crescenzo, E. Aragona, R. Chiesa, R. Castellano, B. Catenaccio, and F. Bandello, "Quantitative optical coherence tomography angiography detects retinal perfusion changes in carotid artery stenosis," *Frontiers Neurosci.*, vol. 15, Apr. 2021, Art. no. 640666.

[9] T. Wang, X. Xu, R. Xiang, J. Wang, and X. Liu, "Association between fundus atherosclerosis and carotid arterial atherosclerosis," *Int. J. Clin. Med.*, vol. 14, no. 5, pp. 282–289, 2023.

[10] J. Liu, J. Wan, W. R. Kwapong, W. Tao, C. Ye, M. Liu, and B. Wu, "Retinal microvasculature and cerebral hemodynamics in patients with internal carotid artery stenosis," *BMC Neurol.*, vol. 22, no. 1, p. 386, Oct. 2022.

[11] A. G. Koutsiaris, V. Batis, G. Liakopoulou, S. V. Tachmitzi, E. T. Detorakis, and E. E. Tsironi, "Optical coherence tomography angiography (OCTA) of the eye: A review on basic principles, advantages, disadvantages and device specifications," *Clin. Hemorheology Microcirculation*, vol. 83, no. 3, pp. 247–271, Apr. 2023.

[12] S. G. Güngör, "Optical coherence tomography angiography in glaucoma," *J. Glaucoma Cataract*, vol. 16, no. 4, p. 171, 2021.

[13] T. R. P. Taylor, M. J. Menten, D. Rueckert, S. Sivaprasad, and A. J. Lotery, "The role of the retinal vasculature in age-related macular degeneration: A spotlight on OCTA," *Eye*, vol. 38, no. 3, pp. 442–449, Feb. 2024.

[14] K. Pradeep, V. Jeyakumar, M. Bhende, A. Shakeel, and S. Mahadevan, "Artificial intelligence and hemodynamic studies in optical coherence tomography angiography for diabetic retinopathy evaluation: A review," *Proc. Inst. Mech. Engineers, Part H, J. Eng. Med.*, vol. 238, no. 1, pp. 3–21, Jan. 2024.

[15] M. Parravano and D. De Geronimo, "Optical coherence tomography angiography in diabetic retinopathy," *Minerva Oftalmologica*, vol. 60, no. 3, Nov. 2018, Art. no. 101206.

[16] O. M. Rifai, S. McGrory, C. B. Robbins, D. S. Grewal, A. Liu, S. Fekrat, and T. J. MacGillivray, "The application of optical coherence tomography angiography in Alzheimer's disease: A systematic review," *Alzheimer's Dementia, Diagnosis, Assessment Disease Monitor*, vol. 13, no. 1, 2021, Art. no. e12149.

[17] L. Wang, S. Shah, C. N. Llaneras, and R. Goldhardt, "Insight into the brain: Application of the retinal microvasculature as a biomarker for cerebrovascular diseases through optical coherence tomography angiography," *Current Ophthalmology Rep.*, vol. 12, no. 1, pp. 1–11, Nov. 2023.

[18] E. L. Esser, L. Lahme, S. Dierse, R. Diener, N. Eter, H. Wiendl, T. Duning, M. Pawlowski, J. Krämer, and M. Alnawaiseh, "Quantitative analysis of retinal perfusion in patients with frontotemporal dementia using optical coherence tomography angiography," *Diagnostics*, vol. 14, no. 2, p. 211, Jan. 2024.

[19] Z. Liu, C. Wang, X. Cai, H. Jiang, and J. Wang, "Discrimination of diabetic retinopathy from optical coherence tomography angiography images using machine learning methods," *IEEE Access*, vol. 9, pp. 51689–51694, 2021.

[20] A. Alfahaid and T. Morris, "An automated age-related macular degeneration classification based on local texture features in optical coherence tomography angiography," in *Proc. Annu. Conf. Med. Image Understand. Anal.*, 2018, pp. 189–200.

[21] A. Alfahaid, T. Morris, T. Cootes, P. A. Keane, H. Khalid, N. Pontikos, P. Sergouniotis, and K. Balaskas, "A hybrid machine learning approach using LBP descriptor and PCA for age-related macular degeneration classification in OCTA images," in *Proc. Annu. Conf. Med. Image Understand. Anal.*, 2019, pp. 231–241.

[22] T.-C. Lin, Y.-C. Jheng, S.-J. Chen, and S.-H. Chiou, "Artificial intelligence machine learning of optical coherence tomography angiography for the diagnosis of age-related macular degeneration," *Investigative Ophthalmology Vis. Sci.*, vol. 61, no. 7, p. 2031, 2020.

[23] M. Heisler, S. Karst, J. Lo, Z. Mammo, T. Yu, S. Warner, D. Maberley, M. F. Beg, E. V. Navajas, and M. V. Sarunic, "Ensemble deep learning for diabetic retinopathy detection using optical coherence tomography angiography," *Translational Vis. Sci. Technol.*, vol. 9, no. 2, p. 20, Apr. 2020.

[24] P. Zang, L. Gao, T. T. Hormel, J. Wang, Q. You, T. S. Hwang, and Y. Jia, "DcardNet: Diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 6, pp. 1859–1870, Jun. 2021.

[25] G. Ryu, K. Lee, D. Park, S. H. Park, and M. Sagong, "A deep learning model for identifying diabetic retinopathy using optical coherence tomography angiography," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, Nov. 2021.

[26] M. Trombini, D. Solarna, G. Moser, and S. Dellepiane, "A goal-driven unsupervised image segmentation method combining graph-based processing and Markov random fields," *Pattern Recognit.*, vol. 134, Feb. 2023, Art. no. 109082.

[27] H. Hu, M. Yao, F. He, and F. Zhang, "Graph neural network via edge convolution for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[28] P. Pradhyumna and G. P. Shreya, "Graph neural network (GNN) in image and video understanding using deep learning for computer vision applications," in *Proc. 2nd Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Aug. 2021, pp. 1183–1189.

[29] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu, "Vision GNN: An image is worth graph of nodes," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Dec. 2022, pp. 8291–8303.

[30] V. Vasudevan, M. Bassenne, M. T. Islam, and L. Xing, "Image classification using graph neural network and multiscale wavelet superpixels," *Pattern Recognit. Lett.*, vol. 166, pp. 89–96, Feb. 2023.

[31] H.-C. Yi, Z.-H. You, D.-S. Huang, and C. K. Kwoh, "Graph representation learning in bioinformatics: Trends, methods and applications," *Briefings Bioinf.*, vol. 23, no. 1, Jan. 2022, Art. no. bbab340.

[32] X. Song, M. Mao, and X. Qian, "Auto-metric graph neural network based on a meta-learning strategy for the diagnosis of Alzheimer's disease," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 3141–3152, Aug. 2021.

[33] S. Zheng, Z. Zhu, Z. Liu, Z. Guo, Y. Liu, Y. Yang, and Y. Zhao, "Multi-modal graph learning for disease prediction," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2207–2216, Sep. 2022.

[34] C. Cui, H. Yang, Y. Wang, S. Zhao, Z. Asad, L. A. Coburn, K. T. Wilson, B. A. Landman, and Y. Huo, "Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: A review," *Prog. Biomed. Eng.*, vol. 5, no. 2, Apr. 2023, Art. no. 022001.

[35] M. Liu, H. Zhang, F. Shi, and D. Shen, "Hierarchical graph convolutional network built by multiscale atlases for brain disorder diagnosis using functional connectivity," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–13, Jun. 2023.

[36] R. Anirudh and J. J. Thiagarajan, "Bootstrapping graph convolutional neural networks for autism spectrum disorder classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3197–3201.

[37] Z. Sun, H. Yin, H. Chen, T. Chen, L. Cui, and F. Yang, "Disease prediction via graph neural networks," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 818–826, Mar. 2021.

[38] C. Mao, L. Yao, and Y. Luo, "MedGCN: Medication recommendation and lab test imputation via graph convolutional networks," *J. Biomed. Informat.*, vol. 127, Mar. 2022, Art. no. 104000.

[39] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1290–1301, Jul. 2022.

[40] G. Lan, Y. Li, M. Hu, Y. Sun, and Y. Zhang, "Knowledge graph integrated graph neural networks for Chinese medical text classification," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, Dec. 2021, pp. 682–687.

[41] J. Hao, J. Liu, E. Pereira, R. Liu, J. Zhang, Y. Zhang, K. Yan, Y. Gong, J. Zheng, J. Zhang, Y. Liu, and Y. Zhao, "Uncertainty-guided graph attention network for parapneumonic effusion diagnosis," *Med. Image Anal.*, vol. 75, Jan. 2022, Art. no. 102217.

[42] Z. Chen, J. Liu, M. Zhu, P. Y. M. Woo, and Y. Yuan, "Instance importance-aware graph convolutional network for 3D medical diagnosis," *Med. Image Anal.*, vol. 78, May 2022, Art. no. 102421.

[43] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, Apr. 2018.

[44] H. Jiang, P. Cao, M. Xu, J. Yang, and O. Zaiane, "Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction," *Comput. Biol. Med.*, vol. 127, Dec. 2020, Art. no. 104096.

[45] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert, "Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease," *Med. Image Anal.*, vol. 48, pp. 117–130, Aug. 2018.

[46] H. Liu, J. Xie, Y. Liu, H. Hao, L. Guo, J. Zhang, and Y. Zhao, "Topology-aware learning for semi-supervised cross-domain retinal artery/vein classification," in *Proc. Comput. Graph. Int. Conf.*, 2022, pp. 41–52.

[47] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976.

[48] H. Hermessi, O. Mourali, and E. Zagrouba, "Multimodal medical image fusion review: Theoretical background and recent advances," *Signal Process.*, vol. 183, Jun. 2021, Art. no. 108036.

[49] T. Zhou, M. Liu, H. Fu, J. Wang, J. Shen, L. Shao, and D. Shen, "Deep multi-modal latent representation learning for automated dementia diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2019, pp. 629–638.

[50] T. Zhou, S. Canu, P. Vera, and S. Ruan, "3D medical multi-modal segmentation network guided by multi-source correlation constraint," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 10243–10250.

[51] N. Zhang, S. Ding, H. Liao, and W. Jia, "Multimodal correlation deep belief networks for multi-view classification," *Int. J. Speech Technol.*, vol. 49, no. 5, pp. 1925–1936, May 2019.

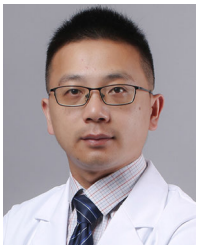
[52] X. Wang, H. Li, Z. Xiao, H. Fu, Y. Zhao, R. Jin, S. Zhang, W. R. Kwapong, Z. Zhang, H. Miao, and J. Liu, "Screening of dementia on octa images via multi-projection consistency and complementarity," in *Proc. Int. Conf. Med. Image Comput.*, 2022, pp. 688–698.

[53] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[54] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lió, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

WENTING LAN received the B.S. degree in clinical medicine science from Wenzhou Medical University, China, in 2007, and the M.S. degree in radiology and nuclear medicine from Zhejiang University, China, in 2018. Since 2007, she has been a Radiologist with the First Affiliated Hospital of Ningbo University, Ningbo, China. Her current research interests include the application of artificial intelligence technology in the diagnosis and construction of prediction models for cerebrovascular and brain tumor diseases.

JINKUI HAO received the B.S. degree in mechanical design and automation from China University of Geosciences, in 2017. He is currently pursuing the Ph.D. degree with the Institute of Biomedical Engineering, University of Chinese Academy of Sciences, Ningbo, China. His research interests include medical image processing and deep learning.



SHENGJUN ZHOU is currently an Associate Chief Physician specializing in neurosurgery. He is also a Visiting Scholar with Toronto West Hospital, Canada. With over ten years of clinical experience, he is highly skilled in surgical and interventional treatments of conditions, such as cerebral aneurysms, cerebral arteriovenous malformations, and arteriovenous fistulas. He also has expertise in procedures related to brain tumors and trauma. He is recognized as one of the pioneering doctors in Ningbo region for performing stroke stent retrieval procedures and has accumulated extensive clinical experience in the surgical treatment of acute large vessel occlusion.

JINGFENG ZHANG received the Ph.D. degree in imaging medicine and nuclear medicine from Huazhong University of Science and Technology (HUST), in 2005. He is currently the Chief Physician with Ningbo No.2 Hospital. He has been engaged in diagnostic imaging and interventional therapy for 28 years. His research interests include diagnostic imaging of thoracic, abdominal, and musculoskeletal diseases.

SHAODONG MA is currently an Assistant Professor with Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China. His research interests include medical image processing and computer-aided diagnosis.

YITIAN ZHAO (Member, IEEE) received the Ph.D. degree in 3-D image analysis from the Department of Computer Science, Aberystwyth University, in 2013. He is currently the Director and a Professor of the Laboratory of Intelligent Medical Imaging (iMED), Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences. His research interests include ophthalmic medical image processing, vessel structure analysis, and eye and brain joint computing. He was the Area Chair of MICCAI, in 2021 and 2022; a Committee Associate Member of IEEE BISP, from 2021 to 2022; the Program Co-Chair of ICIMH, in 2021, and VSIP, in 2021; and a Program Committee Member of AAAI, in 2021, and the MICCAI Workshop.