RESEARCH ARTICLE

Classification of Diabetic Retinopathy Disease Levels by Extracting Topological Features Using Graph Neural Networks

SUMOD SUNDAR[1], (Member, IEEE), AND S. SUMATHY[2]
[1]School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore 632014, India
[2]School of Information Technology and Engineering, Vellore Institute of Technology (VIT), Vellore 632014, India

Corresponding author: S. Sumathy (ssumathy@vit.ac.in)

ABSTRACT Diabetic retinopathy happens due to damage in blood vessels and is the prominent reason for blindness worldwide. Clinical experts observe the fundus images to diagnose the disease, but it is often an error-prone and tedious task. Computer-assisted techniques will help clinicians to detect the disease severity levels. In medical imaging, experiments of automated diagnosis using CNN produce impressive results. Even though disease classification tasks in retinal images via CNN face difficulty in retaining high-quality information at the output. A new deep learning methodology is proposed based on a graph convolutional neural network (GCNN). The proposed model aims to extract the essential retinal image features effectively. The work focuses on extracting the features using a Variational autoencoder and identifying the underlying topological correlations using GCNN. The experiments are carried out using two datasets: Kaggle and EyePACS datasets. The performance of the proposed model is evaluated using accuracy, U-kappa, sensitivity and specificity metrics. The results outperform when compared with other state-of-the-art techniques.

INDEX TERMS Diabetic retinopathy, graph neural networks, variational auto encoders, retinal image classification.

I. INTRODUCTION

For diabetic patients worldwide, diabetic retinopathy (DR) ends in premature blindness. Retinal disorders should be diagnosed and treated using observations of blood vessels. Early diagnosis, fundus screening, and timely intervention can prevent diabetic retinopathy, a severe cause of vision loss that may result in blindness [1]. More precise early screening techniques are required in high-risk groups to reduce the threat of vision loss by retinal disorders. Fundus pictures with expert interpretation are an acceptable screening method for preventing blindness [2]. The primary cause of diabetes's clinical signs is an increase in blood glucose levels, which can harm the retina's blood vessels when present for an extended period. DR causes no visual problems in the early stages. Complications in vision, such as floating patches or dark lines, fuzziness or fluctuation, weak or dead areas of vision,

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente García-Díaz.

and progressive blindness, may also develop as the condition proceeds. Figure 1 illustrates the severity phases of retinal neovascularization, which include non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR), as a result of venous beading. The best approach for protecting the patient's vision is an early diagnosis and timely treatment. Therefore, a considerable requirement exists to prevent lifelong retinal deterioration by efficient evaluation techniques for differentiating retinopathy levels in visual impairment.

Ophthalmologists must put in a lot of time and effort to manually detect diabetic maculopathy. To improve the retinal image structure analysis, image processing methods are needed to design Computer Assisted Diagnostic (CAD) systems. Conventional machine learning-based and deep learning (DL)-based methods exhibit excellent performance while performing automatically grading of DR and DME. Traditional machine learning techniques have the advantage of requiring fewer data and processing resources to train

IEEE *Access*

S. Sundar, S. Sumathy: Classification of DR Disease Levels by Extracting Topological Features Using Graph Neural Networks
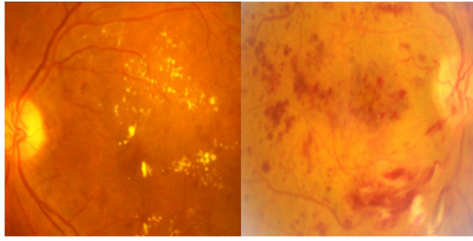

FIGURE 1. DR affected images.

the algorithm. However, for feature engineering, domain knowledge is essential.

Gangwar and Ravi [3] designed a novel CNN model to grade DR fundus images. The model consists of 10 layers of 3×3 convolutions similar to VGGNet and uses leaky ReLu as the activation function. The model aims to detect elements of the blood vessels, such as hemorrhages, exudate, and micro-aneurysms, and then classify them according to whether DR is absent, mild, moderate, severe, or proliferative. The network struggled to acquire deep enough characteristic learning to identify some of the more complicated DR components, as shown by the network's poor sensitivity, especially in the mild and moderate classes. To train a customized convolution network to learn the discriminative features in a colour fundus image for DR detection, Gargeya and Leng [4] applied the technique of deep residual learning as in ResNet. A convolutional visualization layer was added at the network's end to highlight the heatmap-based regions. Even though the network is computationally less expensive, the sensitivity, specificity and AUC values are poor while experimenting with a small MESSIDOR dataset. Cost-effective, robust, and automatic grading without clinical assistance are the key advantages of the proposed system. However, the algorithm needs more optimization for clinical adaptations. Li et al. [5] implemented a multitask algorithm that generated a feature map from retinal images using ResNet50. Attention modules are then used to detect the correlation between two DR grade severity levels. Pires et al. [6] presented a new CNN architecture with convolution and pooling layers similar to VGG-16. The work investigated the network performance in three aspects: using a balanced dataset obtained by data augmentation, multiresolution training, and robust feature-extraction augmentation. Choi et al. [7] developed a deep learning and machine learning-based model for performing multiclass classification of retinal diseases. The deep learning model includes a random forest transfer learning-based VGG-19 architecture with which a 10-class and 3-class classification was performed. Results prove that the deep learning model for 10-class classification is less effective due to a smaller number of images in the STARE dataset. Deep learning techniques were ineffective due to the small size of the dataset. The literature shows that transfer learning techniques produce satisfactory results in image classification tasks [8].

Although deep learning effectively captures underlying patterns in data, there are many applications in which data is represented graphically. Existing methods lack a unified objective for inter- and intra-modality consistency learning and struggle with out-of-sample data. To address these challenges, a Graph Embedding Contrastive Multi-modal Clustering network (GECMC) that integrates representation learning and multi-modal clustering is proposed to enhance both capacities simultaneously [9]. To address the challenges of multi-view learning, the idea of Learnable Graph Convolutional Network and Feature Fusion (LGCN-FF) that incorporates a learnable Graph Convolutional Network (GCN) and a feature fusion network is proposed [10]. The network combines features from various views using multiple sparse autoencoders and a fully-connected network, resulting in a comprehensive representation that captures the characteristics of all views. To tackle the limited interpretability and research gaps in multi-view learning of Graph Convolutional Networks, a novel framework called Interpretable Multi-view Graph Convolutional Network (IMvGCN), which focuses on providing interpretability while solving multi-view learning problems is proposed [11].

To our knowledge, only a few kinds of research have been performed on retinal image classification using graph neural networks. The images can be modelled as a graph structure where each pixel is represented as a node [12]. Deep graph correlation network (DGCN) that utilizes a convolutional graph network can capture natural correlations between independently learnt retinal image features [13]. Graph convolutional networks used for multi-label classification can be used to learn the complicated topology between lesion labels [14]. The vessel graph network (VGN) proposed by [15] used a graph neural network (GNN) to transfer information along vessel structures, have extracted hierarchical patterns in an image. However, works of GNN models that exploit the global structure of vessel shape and their local appearances are limited.

This work consists of the following contributions:
- A Hybrid Graph Convolutional Network (HGCN) that integrates Variational Autoencoder and GCN into a single design.
- GCN layers are used to generate GCN features to take advantage of these discriminative features while learning graph representations.

II. MATERIALS AND METHODS
The architecture of the proposed method to classify retinal images based on their DR severity measures is given in Figure 2.

A. ROI EXTRACTION USING FCN
A Fully convolutional network will map pixels of an image into its pixels using a convolutional neural network [16]. The input RGB retinal image is of dimension 224×224, and the output produced is a 224×224 binary image. The FCN model is constructed by stacking convolution

S. Sundar, S. Sumathy: Classification of DR Disease Levels by Extracting Topological Features Using Graph Neural Networks
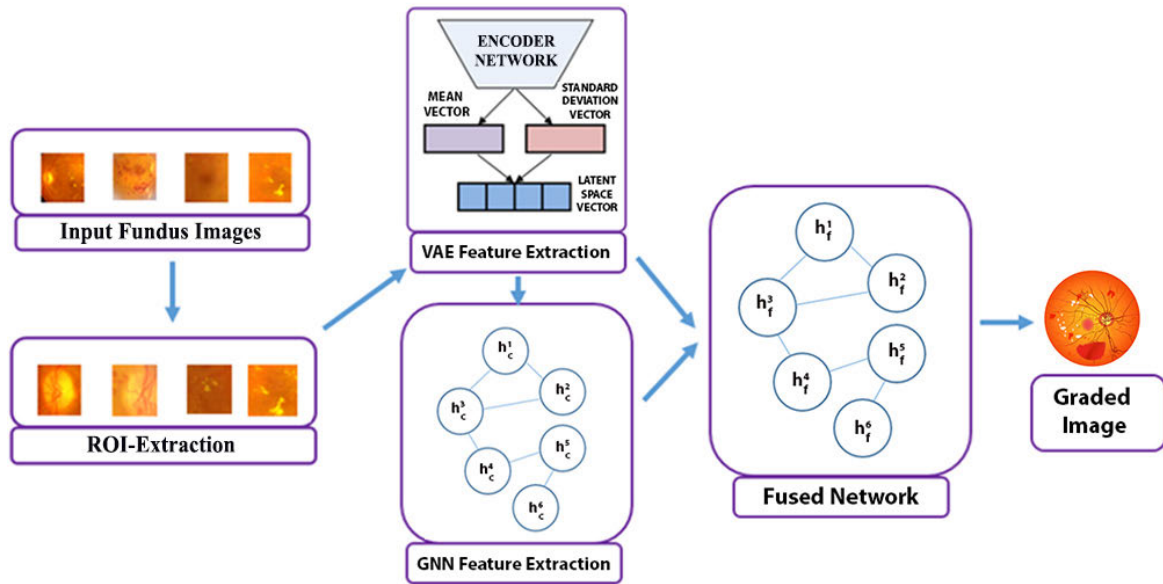
IEEE *Access*

FIGURE 2. Architecture of the proposed method.

blocks composed of 2D convolution layers (Conv2D) and regularizers such as Dropout and BatchNormalization. The VGG-16 architecture consists of 13 convolution layers, five pooling layers and three fully connected layers, which total 21 layers, but it has only 16 weight layers. The kernel size of convolution layers is 3, where the stride and padding values are 1. The kernel size and the pooling stride value of pooling layers are 2. The first and second fully connected layers use "RELU" as the activation function, whereas the third uses the "Softmax" function. The number of channels is then converted into the number of classes using a convolutional layer, and the height and width of the feature maps are modified to match those of the input image using the transposed convolution process. Hence, the output image has the exact dimensions as the input image. The output channel comprises the classes predicted for the input pixel in the exact spatial location as the input pixel. The image features are extracted by utilizing a ResNet-18 model trained on the ImageNet dataset. It is then fine-tuned using the corresponding datasets.

B. VAE FEATURE EXTRACTION

The extracted region proposals from the retinal fundus images using FCN are given into the variational auto-encoder (VAE) to extract the features. VAE is a generative model that analyzes the training data's probability function. The data distribution's mean and standard deviation are estimated using a mean vector and a standard deviation vector, respectively. VAE is utilized for feature extraction since auto-encoders feed input into encoded vectors whose latent space can map data available in the continuous form [1], [17].

The encoder module encodes the images into a latent representation space z. The encoder can be represented as input data x generates output z over the parameter. This input image is given to the encoder, which outputs two latent variables and the distribution parameters. After quantizing the pixel data, the encoder module uses a Gaussian distribution model, and noisy values are minimized by producing a histogram of the final image. Kullback-Leibler (KL) divergence approximates the VAE encoder module's loss function Li. It illustrates how this normal distribution's parameters differ from a unit normal distribution. This divergence is a regularizer added between the encoder's distributions and p(z). A function Li that estimates the quantity of information lost to represent z using q.

$$L_i = KL\left(q_\theta\left(z \mid x\right) \mid p\left(z\right)\right) \tag{1}$$

This function aimed to minimize the divergence magnitude by optimizing the probability distribution parameters μ and σ The more it optimizes, the parameters tune the output to appear similar to the target data distribution. The difference is calculated by KL loss over the distribution can be represented with components as:

$$\sum_{i=}^{n} \sigma_i^2 + \mu_i^2 - \log\left(\sigma_i\right) - 1 \tag{2}$$

The ROI extracted from the previous phase is compacted using VAE and generates latent feature representation. This feature matrix of dimension 330×220 was extracted from all grading levels of the image. The ROI from the earlier step is compressed using VAE to create a latent feature representation. The 330×220 feature matrix was built using

IEEE Access

S. Sundar, S. Sumathy: Classification of DR Disease Levels by Extracting Topological Features Using Graph Neural Networks

the grading levels of the full image. The decoder part of the VAE is removed after training and extracting features from the fundus images. On the next classification level, these trained feature parameters are fed into the GCN classifier model for accurate grading of retinal disease.

C. GCN FEATURE EXTRACTION

Graph Convolutional Networks (GCN) use an image's node feature information and graph. Similar to the convolutional operation of a conventional CNN on an image, spatial-based approaches define graph convolutions based on a node's spatial relations. Images can be treated as a particular type of graph, where each pixel represents a node. A 3×3 patch is filtered after a weighted average of the central node's and its neighbors' pixel values across each channel is computed. A graph neural network model can be defined as f(Mat, X) and the rule for layer-wise propagation can be defined as:

$$Act^{(l+1)}\sigma = \left(\tilde{D}^{-\frac{1}{2}} \tilde{X} \tilde{D}^{-\frac{1}{2}} Act^{(l)} W^{(l)} \right) \qquad (3)$$

Here, $\tilde{X} = X + I_M$ is the adjacency matrix of the undirected graph G with added self-connections. I_N is the identity matrix, $\tilde{D}_{ii} = \sum_j \tilde{X}_{ij \, and} \, W^{(l)}$ is a layer-specific trainable weight matrix. $\sigma(\cdot)$ denotes an activation function, such as the ReLU $(\cdot) = \max(0, \cdot) \cdot H^{(l)} \in R^{M \times D}$ is the matrix of activations in the lth layer; $Act^{(0)} = $ Mat.

A neural network model can be created by stacking numerous convolutional layers on top of one another. A stacked model can lessen the overfitting problem of local neighborhood structures in graphs. Deeper models can be built and features extracted using this layer-by-layer linear computation. The convolution operation on the GCN can be formulated as:

$$\theta_0' x - \theta_1' D^{-\frac{1}{2}} X D^{-\frac{1}{2}} x \qquad (4)$$

where θ_0' and θ_1' are the parameters. The entire graph may share the filtering parameters. The filters can be further applied to convolve the kth-order neighborhood of a node, where k is the number of successive filtering operations or convolutional layers in the neural network model.

Let's define a signal as Mat $\in R^{M \times C}$ with C input channels, where the C-dimensional feature vector is framed for every node and F filters or feature maps as:

$$Z = \tilde{D}^{-\frac{1}{2}} X \tilde{D}^{-\frac{1}{2}} \text{Mat} \Theta \qquad (5)$$

where $\Theta \in R^{C \times F}$, F is the filter parameter matrix and $Z \in R^{M \times F}$ is the convolved signal matrix. As a result, implementing this filtering operation as the product of two matrices—one dense and one sparse—can be done effectively. The output of the last layer $Z = \{Z_1, Z_2, \ldots, Z_k\}$, $Z \in R^{M \times F}$ is the graph representation, and the feature matrix obtained is notated as hg. The entire process involved in the proposed model is summarized in algorithm 1.

Algorithm 1 Fused Graph Convolutional Neural Network

Input: Fundus Images (X, Y); where Y = {y/y ∈ {Normal, Mild, Moderate, Severe, PDR}}
Output: Model to classify the Retinal image x ∈ X
Extract ROI using Fully Convolutional Neural Network Segmentation
    Design FC Layer 1
        Stack 2D Convolutional Layers (filters=64, kernel_size=1, strides=1)
        Apply Dropout, Batch Normalization blocks,
        Apply the RELU Activation function
    Design FC Layer 2
        Stack 2D Convolutional Layers (filters=len_classes, kernel_size=1, strides=1)
        Apply Dropout, Batch Normalization blocks,
        Perform GlobalMaxPooling
        Add 1×1 Conv Layer
    Perform Transposed Convolution
    Assign crop size, Loss= SoftmaxCrossEntropyLoss
    Train and extract pixels
Extract Features using Variational Auto Encoder (VAE)
    Design Encoder
        Flatten Input Image
        Define Dense Layers (200)
    Design Decoder
        Define Dense Layers (200)
    Wrap Encoder and Decoder units
    Train and re-estimate Gradient Descent
        $\nabla \theta, \phi Eq\phi(z)[\log p\theta(x,z) - \log q\phi(z)]$
        where observed x ∈ X where X can be continuous or discrete, and latent z ∈ Rk
        Freeze Decoder and extract feature from encoder
Graph Learning (GL)
    Perform Convolution operation

$$\theta_0' x - \theta_1' D^{-\frac{1}{2}} X D^{-\frac{1}{2}} x$$

Fuse features from VAE and GL: h = concatenate [hc, hg]
Apply Softmax Layers

D. ALGORITHM

The first step is Region of Interest segmentation, using a Fully Convolutional Neural Network (FCN). The FCN is designed with two Fully Convolutional layers. Each FC layer is followed by 2D convolutional layers with 64 filters and a kernel size of 1. Dropout and batch normalization blocks are applied to prevent overfitting, and the rectified linear unit (ReLU) activation function is used to introduce non-linearity. The second FC layer has a number of filters equal to the number of classes to enable classification. The FCN helps identify and extract the relevant regions from the retinal images for further analysis. After segmenting the ROIs, the algorithm employs a Variational Auto Encoder (VAE) to extract features from these regions. The VAE consists of an encoder and a decoder. The encoder takes the segmented

S. Sundar, S. Sumathy: Classification of DR Disease Levels by Extracting Topological Features Using Graph Neural Networks

IEEE*Access*

ROI as input and flattens it. Dense layers with 200 units are used to capture the latent representation of the image. The decoder then reconstructs the image from the latent space. The VAE is trained using gradient descent to optimize the reconstruction loss and the Kullback-Leibler (KL) divergence between the prior and posterior distributions of the latent space. This process allows the VAE to learn meaningful representations of the retinal images. Once the VAE is trained, the decoder is frozen, and the encoder is used to extract features from the ROIs. These features capture the distinctive characteristics of the retinal images, which are crucial for accurate classification. Graph Learning (GL) is performed by applying convolutional operations to the features extracted from the VAE. This step aims to enhance the representation of the features by incorporating graph-based relationships. This can help capture contextual information and further improve the classification performance. The features obtained from both the VAE and GL are then fused by concatenating them, resulting in a combined feature vector. Finally, a softmax layer is applied to the fused features to classify the retinal image into the appropriate severity category. The softmax layer provides the probability distribution over the different severity classes, allowing the model to make predictions.

E. EXPERIMENTATION ON DATASETS

The proposed model was initially trained using the Kaggle dataset (dataset 1) containing 3464 high-quality retinal images. The dataset included high-resolution images acquired under various imaging settings. Dataset 2, used for experimenting with the proposed model, contains 35000 high-resolution fundus images provided by EyePACS for diabetic competition. These images are captured under various imaging conditions. Contrast enhancement was performed on both datasets to adjust the image's bright or dark pixels to extract its hidden features. The contrast between the retinal background and blood vessels in fundus images is very low. Due to the imbalance issue, augmentation is performed on both datasets, and the data size is increased. In the course of model training, an iterative procedure is employed to generate data for every mini-batch. Techniques namely horizontal flip, width shift, height shift, fill mode, and zoom range are used to increase the count of images. The parameters which are used for augmenting retinal images are shown in table 1. The number of images present in both datasets before and after augmentation is shown in table 2.

TABLE 1. Parameters and values used for image augmentation.

Parameter	Values
Zoom Range	0.1
Rotation range	15
Fill mode	Nearest
Horizontal flip	True
Width shift	0.1
Height shift	0.1

The ground truths corresponding to the images were available in both dataset repositories.

TABLE 2. Dataset description.

| Abnormalities	Kaggle Dataset (Dataset 1)		EyePACS Dataset (Dataset 2)	
	No. of samples before augmentation	No. of samples after augmentation	No. of samples before augmentation	No. of samples after augmentation
Normal	-	-	25810	25810
Microaneurysms (MA)	1800	1800	2443	25810
Soft Exudates (SE)	370	1800	5292	25810
Hard Exudates (SE)	999	1800	873	25810
Hemorrhages (HE)	295	1800	708	25810
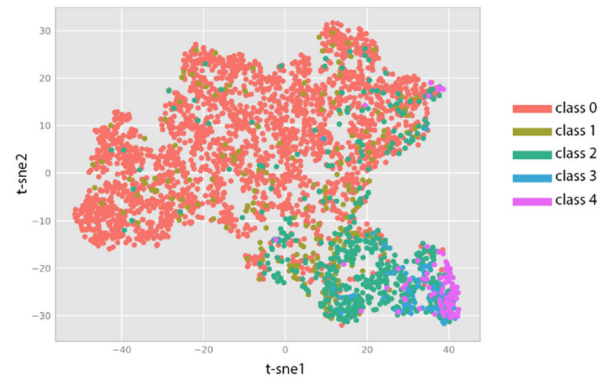
FIGURE 3. Latent vector space representation using t-SNE.

The distribution of severity class obtained on EyePACS dataset is illustrated in figure 3 using t-SNE. We can observe that classes 0, 2, and 4 are distinctly separated in both spaces. However, class 0 and class 1 are not distinctly separated. In the case of classes 3 and 4, although the separation is not perfect, it is possible to identify a discernible difference in the location of both classes for both spaces.

III. RESULTS AND DISCUSSIONS

To classify diseases using computers, the performance of experiments on retinal lesions must be evaluated. As a multiclass task, retinal fundus image classification could be classified as binary, multi-categorical classification, or ordinal regression. The proposed model is viewed as a multi-categorical classification model. The proposed method for classifying DR images has been experimented with, and the sensitivity, specificity, accuracy and U-kappa scores are evaluated. The kappa coefficient measures the agreement score when classifying data into mutually exclusive categories. Galton [18] and Smeeton and Nigel [19] initially used it in

their works. It is estimated as:

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \qquad (6)$$

where P_o indicates the class observed and Pe is the actual class of the given image. The definitions of the various U-kappa score ranges are displayed in Table 3. For the proposed approach, a moderate U-kappa value is attained.

TABLE 3. Definition of U-Kappa score.

Range of U-kappa score	Concordance
Negative	Poor
0.01–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1	Almost perfect

For a simpler model building with eager execution, the operations are carried out using the Keras API in the TensorFlow platform created by the Google Brain Team. Training is performed in the "standard_gpu" configuration powered by the NVIDIA Tesla V100 GPU unit on PowerEdge R740 server. The NVIDIA Tesla V100 is a high-end graphics processing unit (GPU) designed based on NVIDIA's Volta architecture. The machine is configured with 640 Tensor cores, 5120 CUDA cores and is designed to deliver high performance for Artificial Intelligence and high-performance computing (HPC) workloads. The model was trained for 2 hours 5 mins, and in total used 6.5e16 FLOPs during training. Then, features are extracted using "/gpu:0," a GPU with a straightforward configuration and a single virtual machine (VM) specification. The proposed technique outperformed other techniques in terms of accuracy when grading DR-affected images.

The proposed method employs both isolated background pixels and vessel pixels as nodes in a graph. To examine the impact of the isolated nodes, they are eliminated them from the graph representation and perform experiments utilizing only the vessel pixels as graph nodes. Each graph node contains CNN features and hidden topological features are shown in figure 4.

A. PROMINENT RELATED WORKS

Some of the prominent works proposed automated methods for detecting diabetic retinopathy through a series of stages including preprocessing, identifying and removing the optic disc, separating and eliminating blood vessels, removing fovea, and extracting features for Micro-aneurysm, retinal hemorrhage, and exudates classification.

Reference [20] used HSI conversion, DE noising in the pre-processing stage and Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to ensure that the illumination is evenly distributed. The technique of Circular Hough Transform (CHT) is used for Optic disc-based detection and removal whereas as a Bias Corrected Separated Possibilistic
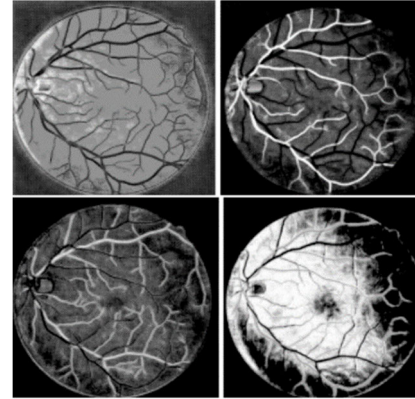

FIGURE 4. Feature maps extracted after GNN module.

Neighborhood FCM (BCSPNFCM) algorithm is introduced for Blood vessel-based segmentation and removal. Fovea elimination is accomplished using morphology dilation whilst splat properties and texture features are fed into Deep Convolutional Networks (DCNN) for classification. In the work of Gharaibeh et al. [21], Normalization, intensity conversion, de-noising using logic statistics with wiener-2 function, and contrast enhancement using CLAHE is used in the preprocessing stage and a modified canny edge detection algorithm is adapted for localization and elimination for optic disc. After segmenting and eliminating blood vessel using Modified spatial weighted fuzzy c-means algorithm, efficient Haralick features are selected using Unsupervised Particle Swarm Optimization based Relative Reduct (US-PSO-RR) algorithm is then given to Maximum Likelihood Classifier (MLC) and Support Vector Mahcine(SVM) for classification. HSI conversion and DE noising is performed by before CHT for optic disc detection, is then followed by Spatially Constrained Possibilistic Fuzzy C-Means (SCPFCM) algorithm for blood vessel segmentation [22]. The SVM-Genetic algorithm (SVMGA) classifier separated the features extracted using Deep Belief Networks (DBN).

B. COMPARISON WITH OTHER MODELS

The performance of the proposed model is compared with other models that use same Kaggle Dataset and EyePACS dataset used for DR and DME grading tasks in the literature.

Samanta et al. [23]: used a transfer learning-based CNN architecture and investigated the classification performance on a smaller dataset. The contrast difference between blood vessels and retinal background is relatively fewer. Hence, contrast enhancement is used as the pre-processing technique to fine-tune the bright and dark pixels. This helped to amplify the pixel contrast of pixels near the retinal area to extract the hidden features. The stacked convolutional layer model is fine-tuned and tested using Inceptionv1, Inceptionv2, Inceptionv3, Xception, VGG16, ResNet-50, DenseNet and AlexNet. Since DenseNet exhibited the best results and simplest architecture among others, it is taken as the baseline architecture. DenseNet121's fully connected

S. Sundar, S. Sumathy: Classification of DR Disease Levels by Extracting Topological Features Using Graph Neural Networks

IEEE *Access*

layers were eliminated and replaced with two fully connected layers of 1024 and a dropout of 0.5. The model has been trained on pre-processed RGB images with dimensions of 360×360 over 50 iterations. Nesterov momentum was chosen as the optimizer with a momentum rate of 0.01. For the first 30 epochs, a learning rate of 0.003 was chosen with a decay of 0.02, is then reduced to 0.001 after 30 iterations.

Reference [24]: The input image size is reduced from 3888×2951 to 786×512. Ensemble models such as Bagging, Boosting and Stacking are used to enhance the model's performance, out of which the Stacking exhibited the best results. Five deep CNN models were ensembled in the method proposed: Resnet50, Inceptionv3, Xception, Dense-121, and Dense169. Categorical cross entropy and Nesterov-accelerated Adaptive Moment Estimation are used as loss functions and optimizers, respectively. The experiment is carried out for 50 epochs with an initial rate α of 10^{-2} and decreased by a factor of 0.1 to 10^{-5}.

Reference [25]: Based on Inception V3, Qummar et al. developed a unique Siamese-like CNN model with weight-sharing layers. All the images are clipped to 299×299 pixels to unify the size of the whole image. Each pixel value in an image is subtracted from the weighted mean of the pixels around it and then added to 50% grayscale. The weights from the Inception V3 model, which was pre-trained on the ImageNet data set, are taken initially. Adam is used as the optimizer, and training is carried out in the server with NVIDIA GeForce GTX1080TI graphics cards.

[26]: Pao et al. resized the whole image into 100×100 pixels, and the green component is extracted from the retinal image. The green component of the fundus photograph was used to compute the entropy image, which was then proposed. Before calculating the entropy images, pre-processing is done using image enhancement by unsharp masking (UM). A bichannel CNN, including the features of both the entropy images of the grey level and the green component, is used.

In our experiments, the entire dataset is divided into 80% and 20% for training and validation respectively. The accuracy and kappa score obtained after the experimentation of the proposed model on Kaggle dataset is compared with the Densenet technique used in [23]. An improvement of 6.59% and 4.19% is obtained for accuracy and kappa score. The results obtained are illustrated in fig. 5. While performing the experiments on EyePACS dataset, the proposed model obtained an accuracy of 90.34%. It is found that the result exhibited 2.85% accuracy improvement than the top models worked on EyePACS dataset classification. Also, the sensitivity and specificity scores obtained were 97.54% and 89.56% for sensitivity and specificity respectively. The results were pretty impressive when comparing the results of other top-performed algorithms [24], [25], [26] experimented on the same dataset. The results are illustrated in fig. 6.

The confusion matrix obtained after performing experiments using datasets 1 and 2 are shown in fig. 7 and fig. 8,
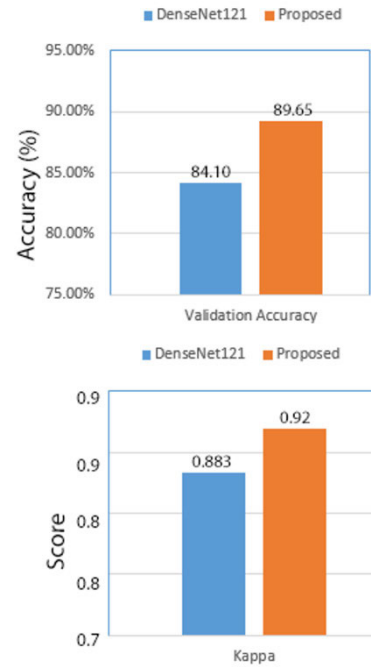

FIGURE 5. Accuracy score comparison and Kappa score comparison of the proposed model with DenseNet121 [23] using Kaggle Dataset.
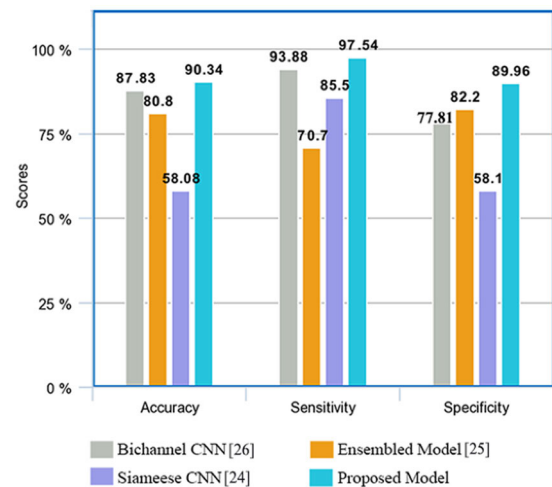

FIGURE 6. Performance comparison of the proposed model with other models using EyePACS dataset.

respectively. The Siamese networks used in [24] require a large amount of data since it works on pair of classes. Also, it is very sensitive to variations in the input. The ensemble model [25] tried to utilize and fuse the features generated using various models. Bi-channel model [26] requires more trainable parameters, thus making the model complex and failing to extract the neighborhood information in an efficient manner.

The accuracy scores observed over various number of iterations after experiments on Kaggle dataset and EyePACS dataset are shown in fig 9 and fig 10 respectively. Kaggle dataset got a convergence in results upon reaching 60 iterations, wherein EyePACS dataset took 150 iterations.
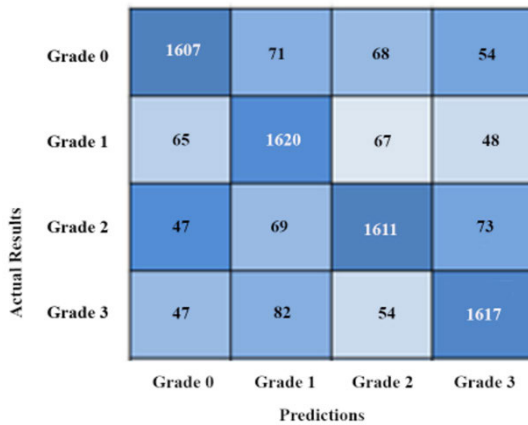
IEEE *Access*

S. Sundar, S. Sumathy: Classification of DR Disease Levels by Extracting Topological Features Using Graph Neural Networks

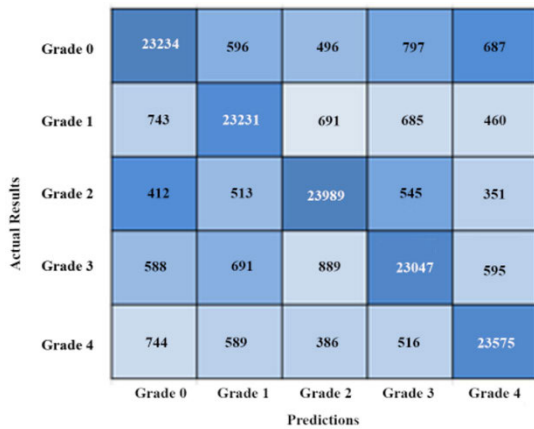**FIGURE 7.** Confusion matrix on grading DR images using Kaggle dataset.

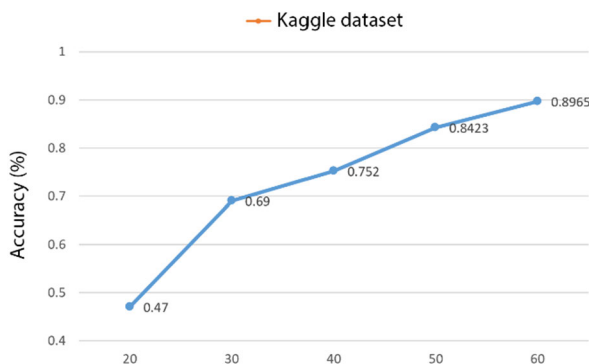**FIGURE 8.** Confusion matrix on grading DR images using EyePACS dataset.

FIGURE 9. Accuracy obtained over various iterations using Kaggle dataset.

C. PERFORMANCE ON FEWER DATA

The images in both datasets are separated at random into percentage groupings of 10%, 20%, 30%, and 40%. Training with these percentages of data generates a model, which is subsequently used for fine-tuning and testing with IDRiD data. The average accuracy is estimated by repeating the
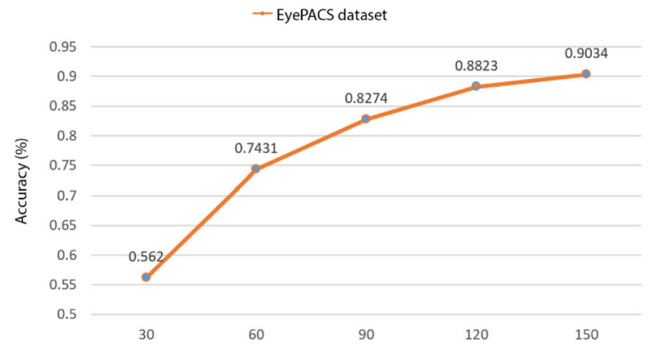

FIGURE 10. Accuracy obtained over various iterations using EyePACS dataset.

training phase for 60 epochs. The performance of the proposed model over various percentages of training data on both datasets is shown in fig. 11. The performance of the model over various folds considering the 80% percentage of training data in Kaggle dataset and EyePACS datasets, respectively. The results are shown in fig. 12. The accuracy of the results was observed to improve as the amount of data utilized for training and no. of folds grew larger, implying that important features were being extracted.
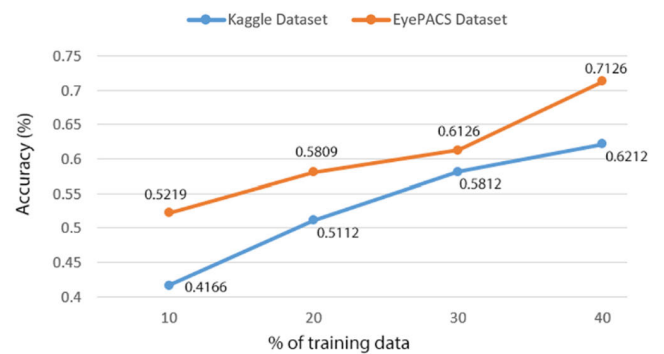

FIGURE 11. Performance of the proposed model on various percentage of data on Kaggle dataset and EyePACS dataset.
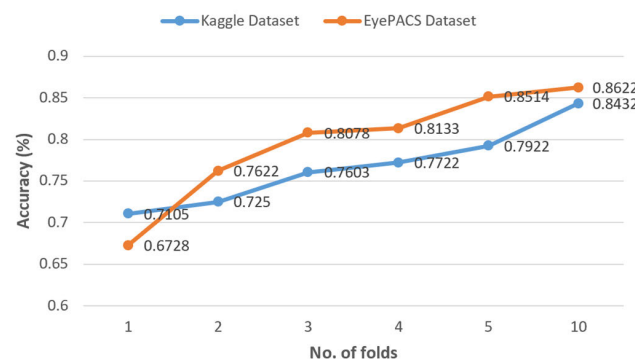

FIGURE 12. Performance of the proposed model over various folds on Kaggle dataset and EyePACS dataset.

S. Sundar, S. Sumathy: Classification of DR Disease Levels by Extracting Topological Features Using Graph Neural Networks

IEEE *Access*

D. PERFORMANCE IN DEALING WITH UNLABELED DATA

To evaluate the capability of the model to learn data distributions in addition to learning of labels, the following experimental methodologies were conducted in two random conditions. The EyePACS dataset is randomly divided into three groups. Scenario 1 contains 50%, 30%, and 20% of the total samples, and scenario 2 includes 65%, 15%, and 20%. In both conditions, the first set comprises data which are labelled, whereas data in the second and third sets are unlabeled. Initial training is conducted on the first set of data, and the trained model is then utilized to label the second set. The second set's annotated labels were not used for training initially; hence the second set is then utilized for testing. Further, testing is conducted on the unlabeled third set using trained model. In both scenarios, the experiment is carried out for 150 epochs, with the weight sets being automatically reinitialized at each stage and technique of early stopping is used. Figure 13 shows the classification accuracies obtained using the model after these observations. Previous sections include the experimental outcomes and performance evaluation.
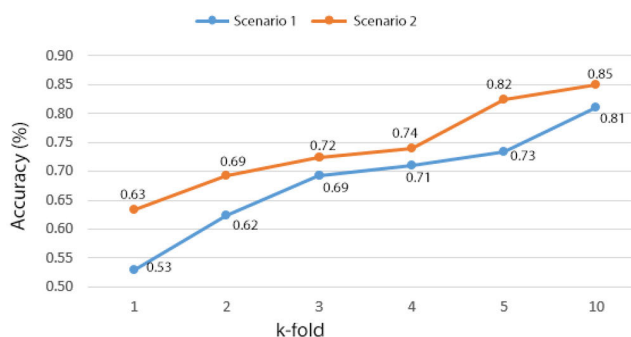

FIGURE 13. Performance of the proposed model in dealing unlabeled data over various folds of unlabeled data using EyePACS dataset.

IV. CONCLUSION

The challenge of performing early analysis of DR using manual techniques raised the necessity for computer-assisted diagnosis. In this work, a fusion method based on GCN is proposed and implemented to classify retinal images according to their DR severity based on their abnormalities. It combines the image features extracted from the variational autoencoder and graph representative features extracted from the graph CNN. The technique learns both local appearances and global vessel structures of the retinal images jointly. Graph convolutional neural networks extract the topological correlations between neighbor pixels using the graph representation and convolution operations. Analyzing and using correlations between each image's appearance features will help to classify retinal severity. This helped to retain feature representation invariant even if the severity-affected pixels are available in different parts of the image. Thus, the proposed technique tried to learn independent features from the retinal images. Also, VAE followed a

tendency to ignore features that occupy few pixels. After implementing the technique, the experimental results are validated using sensitivity, specificity, accuracy and U-kappa scores. It outperformed the top performed algorithms used for retinal image's DR severity. An accuracy improvement of 6.59% and 2.85% is obtained on comparing with Densenet and Bichannel CNN that worked on Kaggle and EyePACS datasets respectively. The proposed fused network model utilized modularity-based graph learning and GCN process to learn the structural influence present in retinal image samples. Additionally, it also leverages more discriminative information present in the images. On the other hand, it is found that the Graph model struggles to utilize images sharp edges and fine details. This happened might be because of the diffusion of GNN vertex features.

REFERENCES

[1] S. Sundar and S. Sumathy, "An effective deep learning model for grading abnormalities in retinal fundus images using variational auto-encoders," *Int. J. Imag. Syst. Technol.*, vol. 33, no. 1, pp. 92–107, Jan. 2023.

[2] S. Kumari, P. Venkatesh, N. Tandon, R. Chawla, B. Takkar, and A. Kumar, "Selfie fundus imaging for diabetic retinopathy screening," *Eye*, vol. 36, no. 10, pp. 1988–1993, Oct. 2022.

[3] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," in *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications*, vol. 1. Singapore: Springer, 2021.

[4] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, Jul. 2017.

[5] X. Li, X. Hu, L. Yu, L. Zhu, C. Fu, and P. Heng, "CANet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1483–1493, May 2020.

[6] R. Pires, S. Avila, J. Wainer, E. Valle, M. D. Abramoff, and A. Rocha, "A data-driven approach to referable diabetic retinopathy detection," *Artif. Intell. Med.*, vol. 96, pp. 93–106, May 2019.

[7] J. Y. Choi, T. K. Yoo, J. G. Seo, J. Kwak, T. T. Um, and T. H. Rim, "Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database," *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0187336.

[8] S. Sumod and S. Sumathy, "Transfer learning approach in deep neural networks for uterine fibroid detection," *Int. J. Comput. Sci. Eng.*, vol. 25, pp. 52–63, Jan. 2022.

[9] W. Xia, T. Wang, Q. Gao, M. Yang, and X. Gao, "Graph embedding contrastive multi-modal representation learning for clustering," *IEEE Trans. Image Process.*, vol. 32, pp. 1170–1183, 2023.

[10] Z. Chen, L. Fu, J. Yao, W. Guo, C. Plant, and S. Wang, "Learnable graph convolutional network and feature fusion for multi-view learning," *Inf. Fusion*, vol. 95, pp. 109–119, Jul. 2023.

[11] Z. Wu, X. Lin, Z. Lin, Z. Chen, Y. Bai, and S. Wang, "Interpretable graph convolutional network for multi-view semi-supervised learning," *IEEE Trans. Multimedia*, early access, Mar. 23, 2023, doi: 10.1109/TMM.2023.3260649.

[12] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[13] G. Zhang, B. Sun, Z. Chen, Y. Gao, Z. Zhang, K. Li, and W. Yang, "Diabetic retinopathy grading by deep graph correlation network on retinal images without manual annotations," *Frontiers Med.*, vol. 9, Apr. 2022, Art. no. 872214.

[14] Y. Cheng, M. Ma, X. Li, and Y. Zhou, "Multi-label classification of fundus images based on graph convolutional network," *BMC Med. Informat. Decis. Making*, vol. 21, no. 2, pp. 1–9, Jul. 2021.

[15] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep vessel segmentation by learning graphical connectivity," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101556.

IEEE Access

S. Sundar, S. Sumathy: Classification of DR Disease Levels by Extracting Topological Features Using Graph Neural Networks

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[17] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in β-VAE," 2018, *arXiv:1804.03599*.

[18] F. Galton, *Finger Prints*. London, U.K.: MacMillan, 1892.

[19] N. C. Smeeton, "Early history of the Kappa statistic," *Biometrics*, vol. 41, p. 795, Jan. 1985.

[20] O. M. Al-hazaimeh, A. Abu-Ein, N. Tahat, M. Al-Smadi, and M. Al-Nawashi, "Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images," *Int. J. Online Biomed. Eng. (iJOE)*, vol. 18, no. 13, pp. 131–151, Oct. 2022.

[21] N. Gharaibeh, O. M. Al-hazaimeh, A. Abu-Ein, and K. M. O. Nahar, "A hybrid SVM Naïve-Bayes classifier for bright lesions recognition in eye fundus images," *Int. J. Electr. Eng. Informat.*, vol. 13, no. 3, pp. 530–545, Sep. 2021.

[22] N. Gharaibeh, O. M. Al-Hazaimeh, B. Al-Naami, and K. M. Nahar, "An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images," *Int. J. Signal Imag. Syst. Eng.*, vol. 11, pp. 206–216, Jan. 2018.

[23] A. Samanta, A. Saha, S. C. Satapathy, S. L. Fernandes, and Y.-D. Zhang, "Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset," *Pattern Recognit. Lett.*, vol. 135, pp. 293–298, Jul. 2020.

[24] X. Zeng, H. Chen, Y. Luo, and W. Ye, "Automated diabetic retinopathy detection based on binocular Siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30744–30753, 2019.

[25] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. A. Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019.

[26] S.-I. Pao, H.-Z. Lin, K.-H. Chien, M.-C. Tai, J.-T. Chen, and G.-M. Lin, "Detection of diabetic retinopathy using bichannel convolutional neural network," *J. Ophthalmol.*, vol. 2020, pp. 1–7, Jun. 2020.

SUMOD SUNDAR (Member, IEEE) received the B.Tech. degree in IT from Anna University, Chennai, and the M.Tech. degree in CSE from the TKM College of Engineering, Kollam, Kerala. He has nearly eight years of teaching experience. He is currently a Research Scholar with the School of Computer Science and Engineering, VIT, Vellore. His research interests include medical imaging and cybersecurity using deep learning techniques. He is the former Academic Relations of the IEEE Kerala YP.

S. SUMATHY received the bachelor's degree in electronics and communication engineering from Madras University, and the M.Tech. degree in computer science and engineering and the Ph.D. degree in computer science from the Vellore Institute of Technology (VIT), Vellore, India. She has nearly 25 years of teaching experience. She is currently a Professor with the School of Information Technology, Engineering, VIT. She has published more than 50 articles in reputed journals in national and international level. Her research interests include trust and reliability in wireless networks, cloud, fog and edge computing, machine learning, and data mining.

● ● ●