Diabetic Retinopathy Grading by Deep Graph Correlation Network on Retinal Images Without Manual Annotations

Guanghua Zhang[1,2†], Bin Sun[3†], Zhixian Chen[1], Yuxi Gao[4], Zhaoxia Zhang[3], Keran Li[5*] and Weihua Yang[5*]

[1] Department of Intelligence and Automation, Taiyuan University, Taiyuan, China, [2] Graphics and Imaging Laboratory, University of Girona, Girona, Spain, [3] Shanxi Eye Hospital, Taiyuan, China, [4] Shanxi Finance and Taxation College, Taiyuan, China, [5] The Laboratory of Artificial Intelligence and Bigdata in Ophthalmology, The Affiliated Eye Hospital of Nanjing Medical University, Nanjing, China

Background: Diabetic retinopathy, as a severe public health problem associated with vision loss, should be diagnosed early using an accurate screening tool. While many previous deep learning models have been proposed for this disease, they need sufficient professional annotation data to train the model, requiring more expensive and time-consuming screening skills.

Method: This study aims to economize manual power and proposes a deep graph correlation network (DGCN) to develop automated diabetic retinopathy grading without any professional annotations. DGCN involves the novel deep learning algorithm of a graph convolutional network to exploit inherent correlations from independent retinal image features learned by a convolutional neural network. Three designed loss functions of graph-center, pseudo-contrastive, and transformation-invariant constrain the optimisation and application of the DGCN model in an automated diabetic retinopathy grading task.

Results: To evaluate the DGCN model, this study employed EyePACS-1 and Messidor-2 sets to perform grading results. It achieved an accuracy of 89.9% (91.8%), sensitivity of 88.2% (90.2%), and specificity of 91.3% (93.0%) on EyePACS-1 (Messidor-2) data set with a confidence index of 95% and commendable effectiveness on receiver operating characteristic (ROC) curve and t-SNE plots.

Conclusion: The grading capability of this study is close to that of retina specialists, but superior to that of trained graders, which demonstrates that the proposed DGCN provides an innovative route for automated diabetic retinopathy grading and other computer-aided diagnostic systems.

Keywords: diabetic retinopathy, retinal image classification, graph correlation network, unsupervised learning, automated diagnosis

INTRODUCTION

Diabetic retinopathy, which has become a severe public health problem, is estimated to affect 200 million patients in the next two decades (1–4). Diabetic retinopathy is a severe etiology of vision loss, possibly causing blindness, which can be avoided by early diagnosis, fundus screening, and temporal treatment (4). Fundus images with professional interpretation are a commonly accepted screening strategy for blindness prevention, with superior capability for in-person dilated eye examinations (5). However, the procedures for observing diabetic retinopathy confront challenges by factors correlated to execution, availability of professionals, and substantial financial sustainability (6–9). The automatic grading of diabetic retinopathy appears to have potential advantages in solving these obstacles, such as increasing efficiency, scalability, and coverage of examining procedures, extending applications in developed regions, and improving patient prevention by providing early diagnosis and referral.

To take full advantage of the clinical application of automatic grading, deep learning technology has been employed to develop a grading system for diabetic retinopathy and achieve considerable progress (10–12). Deep learning adopts artificial intelligence and representation learning tools to address large amounts of data and informative features (13, 14). Recent studies (15–17) have elaborated highly accurate sensitivity and specificity (>90%) in the automatic detection of diabetic retinopathy from fundus images, principally involving accurately annotated fundus images from experienced and professional experts. Recently, the performance of deep learning systems in the examination of diabetic retinopathy is close to that of expert-level diagnoses for grading fundus images, with the help of extensive labeled fundus images from accurate grading by professional manpower. In clinical grading, concentrating extensive professional human resources to provide annotations for developing deep learning systems is unreasonable. Therefore, the performance of an ideal automatic system should also have an excellent capability for detecting diabetic retinopathy from fundus images without manual annotations.

This study mainly aimed to explore the automatic diabetic retinopathy grading without requiring manual annotations on fundus images. This is developed from advanced deep learning technology, thereby saving large amounts of expensive and time-consuming professional manpower. Furthermore, extensive validations on two publicly available diabetic retinopathy grading datasets [EyePACS-1 (18) and Messidor-2 (19)], which consist of various severe clinical challenges, demonstrate the accessibility for automatic grading without manual annotations.

The automatic diabetic retinopathy grading system aims to assist in early diagnosis and grading, which is crucial in diabetic retinopathy screening procedures and requires robust discernment and sufficiently wide computer screens. Several representative attempts have achieved promising effectiveness for grading fundus images (1, 7, 8, 10, 19).

Specifically, Abràmoff et al. (20) developed a deep learning enhanced algorithm for the automatic grading of diabetic retinopathy, and it achieved significantly better performance [sensitivity, 96.8%; specificity, 87.0%; and area under the curve (AUC), 0.980% on MESSIDOR-2 dataset] than did previous methods. Gulshan et al. (18) applied a specific type of neural network optimized for image classification using a retrospective development dataset of 128,175 retinal images for automated grading of diabetic retinopathy and diabetic macular oedema in a retinal fundus photograph. This maintained a sensitivity, specificity, and an AUC of 90.3, 98.1, and 0.991% on EyePACS-1 dataset and 87.0, 98.5, and 0.990% on MESSIDOR-2 dataset, respectively. Ting et al. (14) evaluated the performance of a deep learning system for diabetic retinopathy and related eye diseases using 494,661 retinal images, and it obtained a sensitivity of 95.0%, specificity of 89.7%, and an AUC of 0.941 for overall diabetic retinopathy grading. Gulshan et al. (21) generalized the deep learning automated diabetic retinopathy system to a population of Indian patients at Aravind and Sankara Nethralaya, achieving a sensitivity of 88.9 and 92.1%, specificity of 92.2 and 95.2%, and an AUC of 0.963 and 0.980 for the Aravind Eye Hospital and Sankara Eye Hospital datasets, respectively. This demonstrates the feasibility of using an automated grading system to expand screening programs. Kathiresan et al. (22) proposed a synergic deep learning model for the automated grading and classification of fundus diabetic retinopathy images, involving various processes of pre-processing, segmentation, and classification. This exhibited an excellent accuracy of 99.28%, sensitivity of 98.54%, and specificity of 99.38% on the Messidor dataset.

Although previous deep learning grading systems demonstrated excellent performance on publicly available or self-collected fundus image data, they require expensive and time-consuming professional labor to annotate large amounts of retinal images to guide deep learning models. That severely restricts the application ability of the deep learning models in clinical practice, leading to the decrease of detection efficiency, and patients can not get timely diagnosis and delay the condition. As such, an excellent diabetic retinopathy grading system should be able to handle fundus image grading without the need for manual annotations to provide flexibility in clinical applications.

MATERIALS AND METHODS

To support this development, the EyePACS dataset, which was retrospectively collected from the Eye Picture Archive and Communication System (EyePACS) in the United States and three eye hospitals in India (Aravind Eye Hospital, Sankara Nethralaya, and Narayana Nethralaya), provides a large-scale set of high-resolution retinal images obtained under a variety of imaging conditions. EyePACS is a flexible protocol and web-based telemedicine system for diabetic retinopathy screening and collaboration among clinicians. All images from EyePACS were obtained from different models and types of cameras, and were de-identified according to the Health Insurance Portability and Accountability Act Safe Harbor prior to transferring to the study investigators. Ethics review and institutional review board exemptions were obtained using the Quorum Review IRB. To implement the clinical validation, we randomly sampled

10,286 retinal macula-centered images from 5,158 patients using the EyePACS data (23), wherein different models and types of cameras affected the visual appearance of the left vs. right eyes of the patients. Some images are shown as one would see the retina anatomically (macula on the left and optic nerve on the right for the right eye). Others are shown as one would see through a microscope condensing lens (i.e., inverted, as one sees in a typical live eye exam).

Furthermore, this prospective study enrolled a second dataset, Messidor-2 (19), to prove the scalability of DGCN. Messidor-2 collected diabetic retinopathy examinations with two macula-centered eye fundus images from each eye, part of which was provided by the Messidor program partners and others were from the Messidor extension of previously published examinations from Brest University Hospital. In contrast to the original Messidor dataset of 1,200 images, it consists of 1,748 digital retinal color images from 874 patients with diabetes. Patients with diabetes were recruited from the Ophthalmology Department of Brest University Hospital (France) between October 16, 2009 and September 6, 2010. Eye fundi were imaged, without pharmacological dilation, using a Topcon TRC NW6 non-mydriatic fundus camera with a 45-degree field of view. Only macula-centered images were included in the dataset. Messidor-Extension contains 345 examinations (690 images in JPG format). Other descriptions of this dataset can be found in (19).

All images in the clinical validation sets were graded by several ophthalmologists for the presence of diabetic retinopathy. Diabetic retinopathy severity is annotated as none, mild, moderate, severe, or proliferative, according to the International Clinical Retinopathy Scale (24). The severity of each grade is as follows:

- None: no apparent retinopathy (no abnormalities).
- Mild: mild non-proliferative diabetic retinopathy (microaneurysms only).
- Moderate: moderate non-proliferative diabetic retinopathy (more than microaneurysms but less severe non-proliferative diabetic retinopathy).
- Severe: severe non-proliferative diabetic retinopathy (any extensive intraretinal hemorrhages in each of the four quadrants, definite venous beading in 2+ quadrants, prominent IRMA in 1+ quadrant, and no signs of proliferative retinopathy).
- Proliferative: proliferative diabetic retinopathy (one or more neovascularization and vitreous/preretinal hemorrhage).

As introduced in the datasets, the graders were CN-licensed ophthalmologists and had diabetic retinopathy diagnosis experience of at least 5 years. Furthermore, every image was randomly sent to graders with more than three grades for implementing impartial judgment.

Statistical Analysis

This study first reports the sensitivity and specificity results in 95% confidence intervals (CIs), which are then compared to the diagnosis from retina specialists and trained graders. The sensitivity and specificity are defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

where TP denotes the number of truly predicted positive samples, FN is the number of falsely predicted negative samples, TN is the number of truly predicted negative samples, and FP is the number of falsely predicted positive samples. To characterize the sensitivity and specificity of DGCN, this study reports the presence of referable diabetic retinopathy (moderate or worse diabetic retinopathy) on EyePACS and Messidor-2 data sets.

Second, this study elaborates on the receiver operating characteristic (ROC) curve and its AUC to measure the performance of DGCN on diabetic retinopathy grading without manual annotations. The ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system, as its discrimination threshold is varied. The AUC measures the entire two-dimensional area underneath the entire ROC curve. ROC is a probability curve, and AUC represents the degree or measure of separability. A higher AUC indicates that the model is better in distinguishing between patients with and without the disease.

Third, an important speculation for DGCN is the t-distributed stochastic neighbour embedding (t-SNE) (25), a technique for dimension reduction that is particularly appropriate for visualizing high-dimensional features. This metric can synthetically perform feature learning ability for deep learning models, and it is widely used in unsupervised frameworks. In addition, we utilized PyTorch and Python to develop our DGCN model and conduct the experiments.

Development of Deep Learning Algorithm

Deep learning (DL) technology is effective for pattern recognition based on the rapid development of computing power (GPUs), especially for automated diagnosis, such as in previous studies (26–28). In this study, DL involves millions of trainable parameters to achieve the target task using objective functions. It aims to generate the diabetic retinopathy severity for retinal image data from patients, which do not require any manual annotations, rather than requiring large amounts of labeled data as previous time-consuming and cost-expensive DL systems have done (11, 12, 21). These works were developed from an early DL architecture of a convolutional neural network (CNN), which allows learning independent representations from a single image. In addition, an advanced DL framework of the graph convolutional network (GCN) was proposed recently and preferred to deal with graph correlations between image samples. Moreover, implementing diabetic retinopathy grading without annotations should exploit the inherent similarities from retinal images and rely heavily on correlations between their feature representations. Therefore, this study proposes a deep graph correlation network (DGCN) under an unsupervised framework,

a novel deep learning technology for the automated diagnosis of diabetic retinopathy. In this study, the DGCN model was trained using unlabelled retinal images to optimize the parameters, which were randomly initialized and further explored the inherent distance correlations supervised by estimated labels. Specifically, the DGCN proposed in this study first uses a CNN as a backbone network for learning independent feature representations from retinal images conducted in a mini-batch. Then, DGCN constructs a K nearest neighbor graph by computing similarities to choose the top K samples, and forwards them into a GCN, constrained by a graph-centre (gc) loss. Through these operations, DGCN can make similar features closer to each other, and discriminant correlations in unlabelled retinal images are initially learned. Furthermore, the vital step of this study is to identify the grade for each retinal image, and DGCN designs a pairwise correlation estimating module to determine whether a pair of retinal image features belongs to the same grade, which could provide pseudo annotations automatically. Finally, the specific DGCN framework in this study integrates contrastive loss and transform-invariant constraints based on automated learned pseudo labels.

For simplicity, we summarized main symbols in this part, including retinal images $I = \{I_1, I_2, \ldots, I_N\}$, CNN network $r(\cdot; \theta)$, and GCN network $G(F_b, A)$, where F_b is the CNN features in a batch, and A is the adjacent matrix for F_b.

Convolutional Neural Network

The DGCN framework proposed in this study contains a joint feature learning module of a CNN and GCN, optimized by a series of loss functions, to update the network parameters.

Inspired by the excellent feature learning capability of convolutional neural network, we deploy the CNN framework to extract representative information from image pixels. Formally, retinal images are denoted by $I = \{I_1, I_2, \ldots, I_N\}$, and resNet-50 (29) is utilized as the backbone CNN $r(\cdot; \theta)$. The learned feature collection of I can be denoted as follows:

$$F = [F_0, F_1, \ldots, F_N] = [r(I_1, \theta), r(I_2, \theta, \ldots, r(I_N, \theta))] \quad (1)$$

where θ denotes trainable parameters of CNN.

Graph Convolutional Network

To exploit the correlations between retinal images, GCN is adopted to formulate a topological structure and learn graph feature representations due to its successful application in a medical data analysis (30–32), especially for learning inherent correlations among different samples.

In a mini-batch $I_b = [I_1, \cdots, I_k, \cdots, I_B]$ (B is the batch size and $1 \leq k \leq B$), this study first builds graph correlations for feature set $F_b = [F_1, \cdots, F_k, \cdots, F_B]$, where K nearest neighbors are employed to connect to each sample. From this, DGCN connects sample F_k with its K nearest neighbors by formulating an adjacent matrix $A \in \mathbb{R}^{B \times B}$,

$$\mathbf{A}_{kj} = \begin{cases} 0, & \text{if } j \in \mathcal{K}_k \\ 1, & \text{if } j \notin \mathcal{K}_k \end{cases} \quad (2)$$

where A_{kj} denotes the correlation between k-th and j-th samples, and $\mathcal{K}_k$ is a collection of K nearest neighbors to I_k.

From the learned CNN feature F_b and adjacent matrix A in a mini-batch, $G(F_b, A)$ represents the graph correlations from I_b. The GCN in this study consists of one input layer and L hidden layers (12). The graph convolution for retinal image representation is as follows:

$$X^{(l)} = \left[X_1^{(l)}, \ldots, X_k^{(l)}, \ldots, X_B^{(l)} \right]$$
$$\sigma \left(D^{-1/2} A D^{-1/2} X^{(l-1)} W^{(l)} \right) \quad (3)$$

where $l = 0, 1, \ldots, L$, and $D = \text{diag}(d_1, d_2, \ldots, d_B)$, thus denoting diagonal matrix with $d_k = \sum_{j=1}^{B} A_{kj} \in \mathbb{R}^{d_{l-1} \times d_l}$. $X^0 = F_b$ as the input of the GCN, $\sigma(\cdot)$, which represents the activation of $\text{ReLU}(\cdot) = \max(0, \cdot)$, and $X^{(l)} \in \mathbb{R}^{B \times d_l}$ is the output from the l-th layer.

The output of the last layer $Z = \{Z_1, \ldots, Z_k, \ldots, Z_B\} \in \mathbb{R}^{B \times m}$ is the graph representation of the mini-batch I_b, and the parameters of θ and $W = \{W^0, W^1, \ldots, W_{(l)}\}$ need to be trained by several loss functions focusing on unsupervised retinal image classification.

Loss Function and Optimisation

One of the most critical issues of DL is the loss function, which provides directions of training and application. The loss functions of this study are as follows: gc loss, pseudo-contrastive (pc) loss, and transform-invariant loss.

This study aims to explore the correlations between retinal images, and it should make similar samples closer to network training. Thus, DGCN adopts a gc loss on graph representations in each mini-batch.

$$\mathcal{L}_{gc} = \frac{1}{2} \sum_{k=1}^{B} \sum_{j=1}^{B} A_{kj} \| Z_j - Z_k \|^2 \quad (4)$$

where Z_k is the center of K-nearest neighbors.

Furthermore, this study proposes a label estimator to compute pseudo-annotations for each sample. This label estimator is based on correlations between each other and only outputs pairwise annotations. Given a pair of samples I_j and I_k from the mini-batch, if both of them have more shared K nearest neighbors, it is possible that they belong to the same category. Mathematically, given i-th and k-th samples, the annotation estimation is computed as follows:

$$l(k, j) = \begin{cases} 0, & \text{if } \| \mathcal{K}_k \cap \mathcal{K}_j \| \leq \lambda \\ 1, & \text{if } \| \mathcal{K}_k \cap \mathcal{K}_j \| > \lambda \end{cases} \quad (5)$$

where λ is a threshold to balance the number of shared neighbors.

According to Eq. 5, pairwise retinal images have contrastive annotations $l(\cdot, \cdot)$ (0 or 1), where 1 denotes that they belong to the same category and 0 means they are in different classes. Subsequently, this study utilizes a pc loss function on each pair of samples from the mini-batch.

$$\mathcal{L}_{pc} = \frac{1}{2N} \sum_{\forall \, I_k, I_j \in I_b} l\,(k, j)\, d^2 + \left(1 - l\,(k, j)\right) \max(m - d, 0)^2$$

(6)

where d represents the Euclidean distance and m is a margin parameter. This loss constraint guarantees that samples with the same pseudo pairwise label have smaller distances and those with different pseudo-annotations have larger distances.

The third concern of this study is the feature scalability of each sample after CNN and GCN; thus, a transform-invariant loss is employed to ensure that the feature representation remains invariant regardless of undergoing random transformation. Concretely, we assumed a transformation $T\,(\cdot)$ (e.g., random rotation) on each retinal image, thus generating a new image $I'_k = T\,(I_k)$ with equal discriminative information to that of the original retinal image. It is expected that the feature representations from DGCN preserve identical information despite suffering transformations, achieved by a transform-invariant (ti) loss,

$$\mathcal{L}_{ti} = \frac{1}{B} \sum_{k=1}^{B} \| Z_k - Z'_k \|$$

(7)

where Z'_k represents the graph feature of I'_k.

Finally, this study employs an overall loss function to conduct optimisation for a deep GCN as follows:

$$\mathcal{L} = \mathcal{L}_{pc} + \alpha \mathcal{L}_{gc} + \beta \mathcal{L}_{ti}$$

(8)

where α and β are balance parameters for three terms.

Application of Deep Graph Correlation Network Algorithm

After sufficient training for DGCN, it is then applied to retinal image grading. This study employs several gallery images $I^g = \left[I_1^g, \ldots, I_i^g, \ldots, I_n^g\right]$ with correct grading annotations, fed into the network and output graph features $Z^g = \left[Z_1^g, \ldots, Z_i^g, \ldots, Z_n^g\right]$. Given a testing retinal image I_k, the DGCN obtains its annotation by

$$Label_k = \arg\min_c \left(Z_k, Z_c^g\right); \quad 1 \le c \le C$$

(9)

where Z_c^g is the sample from the c-th class. This prediction strategy searches for the smallest distance between I_k and all gallery images, and predicts its annotation using the label of the matched image.

RESULTS

In our experiments, the number of retinal images from EyePACS-1 and SXEye were 10,286 and 1,748, respectively. Each retinal image was graded by ophthalmologists 4–6 times, which were only utilized in the evaluation of this study. The EyePACS-1 data were collected from 5,158 persons, and the grading was classified into five levels: none, mild, moderate, severe, and proliferative. Specifically, EyePACS-1 consists of 2,276

(22.1%) none, 2,038 (19.8%), 1,982 (2161%) severe, and 1,829 (17.8%) proliferative. Data samples are shown in **Figure 1**. The deep GCN is directly trained for these two data sets when there is no need to involve manual annotations. In the evaluation, we prepared two extra data sets as a baseline with an average of 100 images from each class to calculate predictions for each retinal image, according to Eq. 9. These two baseline sets were collected from the EyePACS-1 and Messidor-2 datasets.

Table 1 reports the performance of the accuracy, sensitivity, and specificity of this algorithm. Specifically, DGCN grades EyePACS-1 with 89.9% accuracy, 88.2% sensitivity, and 91.3% specificity and Messidor-2 with 91.8, 90.2, and 93.0%, respectively. Moreover, this study compares DGCN with a retina specialist and one trained grader to conduct a comparison. From the results, DGCN performs distance of 1.8 (91.7–89.9)% accuracy, 1.3 (89.5–88.2)% sensitivity, and 1.9 (93.2–91.3)% specificity to a retinal specialist, using EyePACS as an example. It should be noted that retinal specialists require professional long-term learning and training with prohibitive cost. In contrast, DGCN is superior to short-term trained graders with approximately 1–2% advantage in each metric. From this comparison, the DGCN proposed in this study performs well with an admitted distance to a specialist, and exceeds the diagnostic capability of the trained grader. Importantly, this study does not involve manual annotations in model development, and its performance is somewhat practical for automated diabetic retinopathy diagnosis.

For the diagnosis of referable diabetic retinopathy, this study also reports the ROC curves for EyePACS-1 and Messidor-2 data sets in **Figure 2**, and their t-SNE plots are shown in **Figure 3**. DGCN achieved an AUC of 0.953 and 0.974 on EyePACS-1 and Messidor-2, respectively (**Figures 2A,B**). From the t-SNE performance in **Figure 3**, the referable diabetic retinopathy (navy points) is less misdiagnosed than normal images (deep pink points), as well as sensitivity and specificity (**Table 1**). Compared to the retina specialist and trained graders, the DGCN proposed in this study achieves satisfactory clinical sensitivity and specificity, as shown in **Figures 2**, **3** and **Table 1**. This demonstrates that the DGCN method can perform automated diabetic retinopathy diagnosis without manual annotations, which yields commendable results and economizes expensive well-trained professional labor.

DISCUSSION

The experimental results on publicly available datasets (EyePACS-1 and Messidor-2) demonstrate that automated diabetic retinopathy grading without any professional annotations is achievable, significantly alleviating economic and human resources. The performance of DGCN is close to that of a retina specialist but superior to that of trained graders. The model without training by annotations is encouraged to involve CNN and GCN to explore inherent correlations between retinal images further. It is also proven that this method is effective,
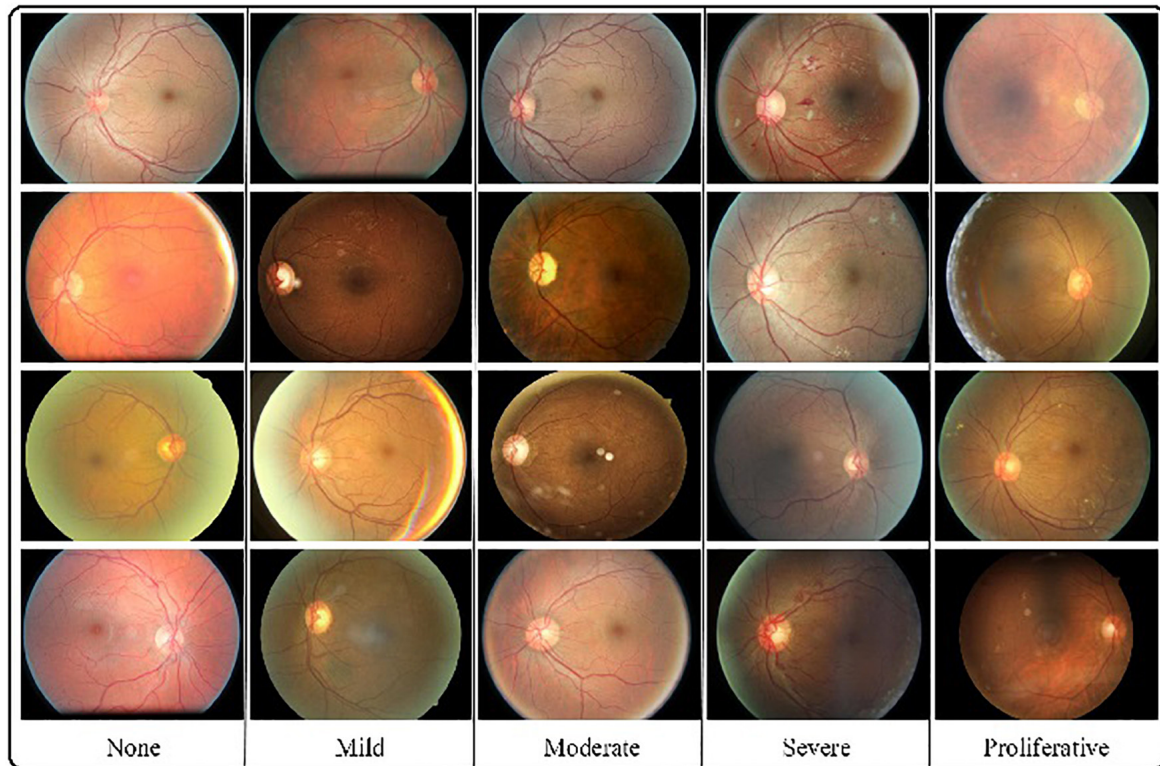
FIGURE 1 | Several examples for each grade. The challenges caused by various viewpoints, illumination, and contrast can be seen in different retinal images.

with an accuracy of 89.9%, sensitivity of 88.2%, and specificity of 91.3% for EyePACS-1, and an accuracy of 91.8%, sensitivity of 90.2, and specificity of 93.0% on Messidor-2.
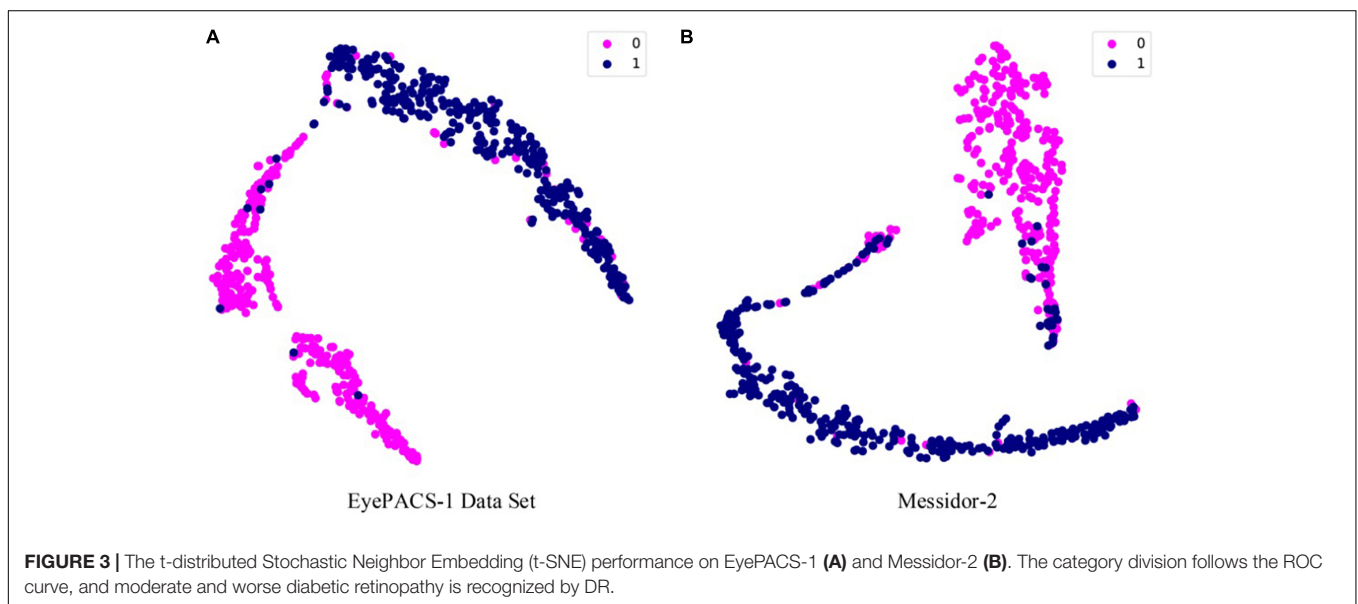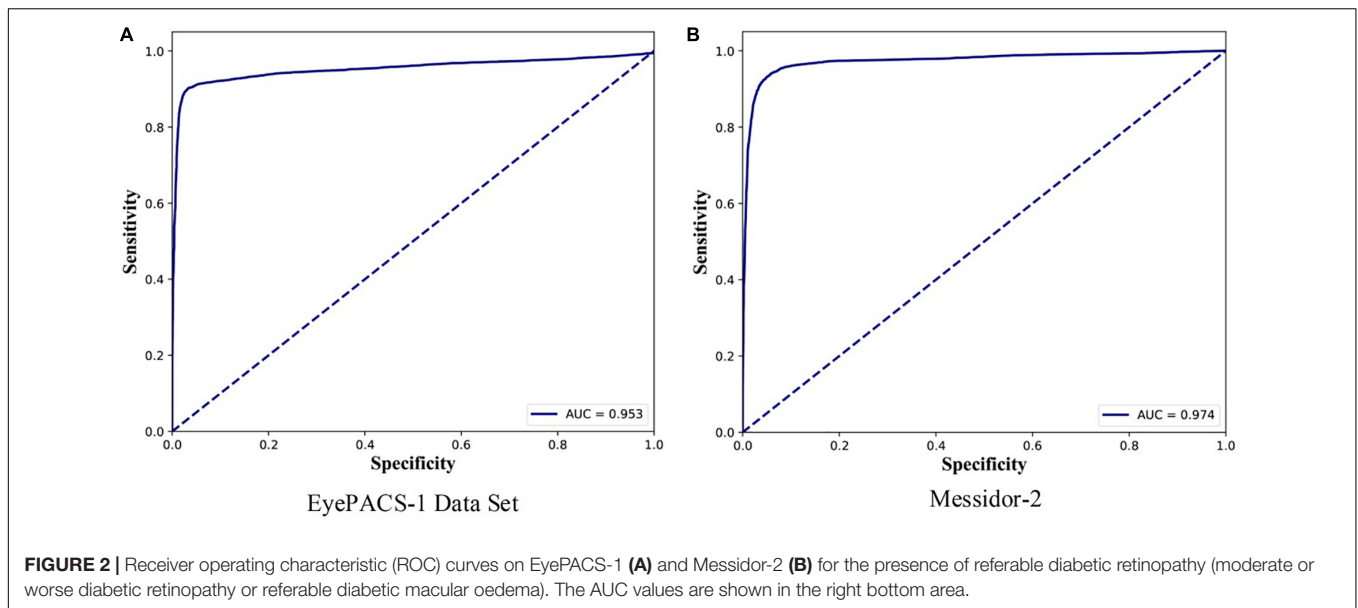
Inspired by successful deep learning applications in biomedical image analysis (33–35), this paper attempts to involve advanced deep learning models into retinal image classification. Considering the challenging variations in illumination, viewpoint, and contrast, retinal image grading is a severe automated diagnosis problem, even with professional annotations. Utilizing correlations between each image appearance feature is of utmost importance, and keeping the feature representation invariant when it suffers from transformations positively affects automated diabetic retinopathy

grading. Aiming at these two points, this study employs a CNN to learn independent features from raw retinal images. They are then fed into a GCN with an established graph from its K nearest neighbors. To draw similar samples closer to each other, the DGCN model attaches a gc loss on each mini-batch to achieve this goal. Furthermore, this study designs a pairwise label estimator for given pairwise retinal features and involves pseudo-labels into a pc loss on each pair of samples. DGCN also realizes the transformation-invariant characteristics of the network by a transformation-invariant loss to preserve identical information. Finally, this study calculates the distances between each sample to a few previous samples. This contains precise annotations and determines the category by annotating the matched sample, which has the smallest distance to the target images. Compared with existing DR grading methods, the advance of our DGCN is that it has excellent capability to automatically learn discriminative information from unlabeled data, which is the easiest data to obtain.

Although the DGCN model performs well in automated diabetic retinopathy diagnosis without the aid of professional annotations, it remains a weakness for models trained by full labels. This method also has limitations on the DR grading task, including redundant training times, guidance of annotated data in applications, and lower sensitivity performance. In general, this study opens a novel economical method for automated disease diagnosis without needing professional manpower. It may produce

TABLE 1 | Performance of sensitivity and specificity of deep graph correlation network (DGCN), compared with a retina specialist and trained graders (95% CI).

Data set	Accuracy	Sensitivity	Specificity
EyePACS-1		**% (95% CI)**	
Retina specialist	91.7 (88.1–93.8)	89.5 (87.2–90.6)	93.2 (88.7–94.8)
Trained grader	88.8 (84.6–92.5)	86.4 (81.2–89.8)	90.4 (87.9–92.7)
DGCN model	89.9 (87.3–91.9)	88.2 (86.4–90.0)	91.3 (89.4–93.3)
Messidor-2		**% (95% CI)**	
Retina specialist	93.3 (89.5–95.8)	91.0 (87.2–93.4)	95.0 (91.7–96.2)
Trained grader	89.7 (86.0–91.8)	87.9 (85.2–90.6)	91.1 (86.9–93.2)
DGCN model	91.8 (89.9–93.2)	90.2 (89.4–91.5)	93.0 (91.5–94.3)

FIGURE 2 | Receiver operating characteristic (ROC) curves on EyePACS-1 **(A)** and Messidor-2 **(B)** for the presence of referable diabetic retinopathy (moderate or worse diabetic retinopathy or referable diabetic macular oedema). The AUC values are shown in the right bottom area.

FIGURE 3 | The t-distributed Stochastic Neighbor Embedding (t-SNE) performance on EyePACS-1 **(A)** and Messidor-2 **(B)**. The category division follows the ROC curve, and moderate and worse diabetic retinopathy is recognized by DR.

more practical performance when utilizing few annotated data and large-scale unlabelled samples, which will be exploited in future studies.

This study elaborates on the effectiveness of deep learning technology in automated diabetic retinopathy diagnosis without manual annotations, rather than fully labeled data, as in previous methods. The feasibility of this research provides a novel ideology for computer-aided systems with economic and easy operability.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

GZ and BS conceived, designed and performed the analysis and wrote the manuscript. ZC contributed to the data and analysis tools and wrote the manuscript. YG performed the analysis and wrote the manuscript. ZZ collected the data and performed the analysis. KL and WY collected the data, wrote the manuscript, and supervised. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Research Funds of Shanxi Transformation and Comprehensive Reform Demonstration

REFERENCES

1. Li X, Dong L, Zhang H. Comparison of fundus fluorescein angiography and fundus photography grading criteria for early diabetic retinopathy. *Int J Ophthalmol.* (2022) 15:261–7. doi: 10.18240/ijo.2022.02.11

2. Li WY, Yang M, Song YN, Luo L, Nie C, Zhang MN. An online diabetic retinopathy screening tool for patients with type 2 diabetes. *Int J Ophthalmol.* (2021) 14:1748–55. doi: 10.18240/ijo.2021.11.15

3. Antonetti DA, Silva PS, Stitt AW. Current understanding of the molecular and cellular pathology of diabetic retinopathy. *Nat Rev Endocrinol.* (2021) 17:195–206. doi: 10.1038/s41574-020-00451-4

4. Roberts-Martínez Aguirre I, Rodríguez-Fernández P, González-Santos J, Aguirre-Juaristi N, Alonso-Santander N, Mielgo-Ayuso J, et al. Exploring the quality of life related to health and vision in a group of patients with diabetic retinopathy. *Healthcare (Basel).* (2022) 10:142. doi: 10.3390/healthcare10010142

5. Rêgo S, Monteiro-Soares M, Dutra-Medeiros M, Soares F, Dias CC, Nunes F. Implementation and evaluation of a mobile retinal image acquisition system for screening diabetic retinopathy: study protocol. *Diabetology.* (2022) 3:1–16. doi: 10.3390/diabetology3010001

6. Abdool Z, Naidoo K, Visser L. Development of a diabetic retinopathy screening model for a district health system in Limpopo Province, South Africa. *Afr Vis Eye Health.* (2022) 81:10. doi: 10.4102/aveh.v80i1.568

7. Chawla S, Chawla A, Chawla R, Jaggi S, Singh D, Trehan S. Trained nurse-operated teleophthalmology screening approach as a cost-effective tool for diabetic retinopathy. *Int J Diabetes Dev Ctries.* (2022) 1–4. doi: 10.1007/s13410-021-01037-0

8. Wan C, Zhou X, You Q, Sun J, Shen J, Zhu S, et al. Retinal image enhancement using cycle-constraint adversarial network. *Front Med (Lausanne).* (2021) 8:2881. doi: 10.3389/fmed.2021.793726

9. Zheng B, Wu M, Zhu S, Zhou H, Hao X, Fei F, et al. Attitudes of medical workers in China toward artificial intelligence in ophthalmology: a comparative survey. *BMC Health Serv Res.* (2021) 21:1067. doi: 10.1186/s12913-021-07044-5

10. Yang WH, Zheng B, Wu MN, Zhu SJ, Fei FQ, Weng M, et al. An evaluation system of fundus photograph-based intelligent diagnostic technology for diabetic retinopathy and applicability for research. *Diabetes Ther.* (2019) 10:1811–22. doi: 10.1007/s13300-019-0652-0

11. Gao Z, Jin K, Yan Y, Liu X, Shi Y, Ge Y, et al. End-to-end diabetic retinopathy grading based on fundus fluorescein angiography images using deep learning. *Graefes Arch Clin Exp Ophthalmol.* (2022) 1–11. doi: 10.1007/s00417-021-05503-7

12. Dai L, Wu L, Li H, Cai C, Wu Q, Kong H, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat Commun.* (2021) 12:1.

13. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539

14. Ting DSW, Cheung CY-L, Lim G, Wei Tan GS, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA.* (2017) 318:2211–23. doi: 10.1001/jama.2017.18152

15. Weng M, Zheng B, Wu M, Zhu SJ, Sun YQ, Liu YF, et al. A preliminary study of a deep learning-assisted diagnostic system with an artificial intelligence for detection of diabetic retinopathy. *Int Eye Sci.* (2018) 18:568–71. doi: 10.3980/j.issn.1672-5123.2018.3.40

16. Niu Y, Gu L, Zhao Y, Lu F. Explainable diabetic retinopathy detection and retinal image generation. *IEEE J Biomed Health Inform.* (2022) 26:44–55. doi: 10.1109/JBHI.2021.3110593

17. Das S, Kharbanda K, Suchetha M, Raman R, Dhas E. Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy. *Biomed Signal Process Control.* (2021) 68:102600. doi: 10.1016/j.bspc.2021.102600

18. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216

19. Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, et al. Feedback on a publicly distributed image database: the Messidor database. *Image Anal Stereol.* (2014) 33:231–4. doi: 10.5566/IAS.1155

20. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci.* (2016) 57:5200–6. doi: 10.1167/iovs.16-19964

21. Gulshan V, Rajan RP, Widner K, Wu D, Wubbels P, Rhodes T, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol.* (2019) 137:987–93. doi: 10.1001/jamaophthalmol.2019.2004

22. Kathiresan S, Sait ARW, Gupta D, Lakshmanaprabu S, Khanna A, Pandey HM. Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. *Pattern Recogn Lett.* (2020) 133:210–6. doi: 10.1016/j.patrec.2020.02.026

23. Eyepacs. *Eyepacs Dataset.* (2020). Available online at: http://www.eyepacs.com/data-analysis (accessed December 18, 2021).

24. American Academy of Ophthalmology. *International Clinical Diabetic Retinopathy Disease Severity Scale Detailed Table.* (2020). Available online at: https://web.archive.org/web/20200110213754id_/http://www.icoph.org:80/dynamic/attachments/resources/diabetic-retinopathy-detail.pdf (accessed December 18, 2021).

25. van der Maaten L, Hinton G. Visualizing data using t-sne. *J Mach Learn Res.* (2008) 9:2579–605.

26. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell.* (2018) 172:1122–31.e9. doi: 10.1016/j.cell.2018.02.010

27. Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. *Nat Commun.* (2020) 11:3543. doi: 10.1038/s41467-020-17280-8

28. Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. *Nat Methods.* (2019) 16:1233–46. doi: 10.1038/s41592-019-0403-1

29. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Las Vegas, NV (2016). p. 770–8.

30. Yu Z, Huang F, Zhao X, Xiao W, Zhang W. Predicting drug-disease associations through layer attention graph convolutional network. *Brief Bioinform.* (2021) 22:bbaa243. doi: 10.1093/bib/bbaa243

31. Zhang YD, Satapathy SC, Guttery DS, Górriz JM, Wang S. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Inf Process Manag.* (2021) 58:102439. doi: 10.1016/j.ipm.2020.102439

32. Chereda H, Bleckmann A, Menck K, Perera-Bel J, Stegmaier P, Auer F, et al. Explaining decisions of graph convolutional neural networks: patient-specific molecular subnetworks responsible for metastasis prediction in breast cancer. *Genome Med.* (2021) 13:1. doi: 10.1186/s13073-021-00845-7

33. Meng Y, Zhang H, Zhao Y, Yang X, Qiao Y, MacCormick IJC, et al. Graph-based region and boundary aggregation for biomedical image segmentation. *IEEE Trans Med Imaging.* (2021) 1–13. doi: 10.1109/TMI.2021.3123567

34. Preston FG, Meng Y, Burgess J, Ferdousi M, Azmi S, Petropoulos IN, et al. Artificial intelligence utilising corneal confocal microscopy for the diagnosis of peripheral neuropathy in diabetes mellitus and prediabetes. *Diabetologia.* (2022) 65:457–66. doi: 10.1007/s00125-021-05617-x

35. Meng Y, Zhang H, Gao D, Zhao Y, Yang X, Qian X, et al. BI-GCN: boundary–aware input-dependent graph convolution network for biomedical image segmentation. In: *Proceedings of The British Machine Vision Conference* (online). (2021).

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