

RESEARCH

Open Access



Multi-label classification of fundus images based on graph convolutional network

Yinlin Cheng^{1,2†}, Mengnan Ma^{1,2†}, Xingyu Li^{2,3} and Yi Zhou^{2,4*}

From International Conference on Health Big Data and Artificial Intelligence 2020 Guangzhou, China. 29 October - 1 November 2020

Abstract

Background: Diabetic Retinopathy (DR) is the most common and serious microvascular complication in the diabetic population. Using computer-aided diagnosis from the fundus images has become a method of detecting retinal diseases, but the detection of multiple lesions is still a difficult point in current research.

Methods: This study proposed a multi-label classification method based on the graph convolutional network (GCN), so as to detect 8 types of fundus lesions in color fundus images. We collected 7459 fundus images (1887 left eyes, 1966 right eyes) from 2282 patients (1283 women, 999 men), and labeled 8 types of lesions, laser scars, drusen, cup disc ratio ($C/D > 0.6$), hemorrhages, retinal arteriosclerosis, microaneurysms, hard exudates and soft exudates. We constructed a specialized corpus of the related fundus lesions. A multi-label classification algorithm for fundus images was proposed based on the corpus, and the collected data were trained.

Results: The average overall F1 Score (OF1) and the average per-class F1 Score (CF1) of the model were 0.808 and 0.792 respectively. The area under the ROC curve (AUC) of our proposed model reached 0.986, 0.954, 0.946, 0.957, 0.952, 0.889, 0.937 and 0.926 for detecting laser scars, drusen, cup disc ratio, hemorrhages, retinal arteriosclerosis, microaneurysms, hard exudates and soft exudates, respectively.

Conclusions: Our results demonstrated that our proposed model can detect a variety of lesions in the color images of the fundus, which lays a foundation for assisting doctors in diagnosis and makes it possible to carry out rapid and efficient large-scale screening of fundus lesions.

Keywords: Diabetic retinopathy, Fundus images, GCN, Multi-label

Background

Diabetic Retinopathy (DR) is the most common and serious microvascular complication in the diabetic population. And it has become the first blinding factor for working-age people worldwide [1, 2]. Timely screening

and treatment of DR have been shown to reduce blindness [3]. In order to reduce the socio-economic burden of vision loss caused by retinal diseases, more accurate early screening procedures are needed in high-risk groups. Common DR diagnostic methods are fundus photography of the retinal and fluorescein fundus angiography (FFA). Fundus photography of the retinal can diagnose patients quickly, but its accuracy depends largely on the experience of the physician. FFA can clearly reflect the pathology of the blood vessels in the retina fundus, but it takes a long time and may cause a variety of adverse

*Correspondence: zhouyi@mail.sysu.edu.cn

[†]Yinlin Cheng and Mengnan Ma have contributed equally to this study

⁴ Key Laboratory of Tropical Disease Control (Sun Yat-sen University), Ministry of Education, No. 74 Zhongshan 2nd Road, Guangzhou 510080, China

Full list of author information is available at the end of the article



reactions [4, 5]. Therefore, compared with FFA, fundus photography of the retinal is often used in the researches of fundus disease diagnosis algorithms, which makes it possible to carry out large-scale fundus disease screening rapidly and efficiently, and is conducive to the early detection and treatment of DR patients [6–14].

The common DR diagnosis method is the retinal fundus images detection. But the fundus images detection is time-consuming and its accuracy also depends on the doctor's experience. With the development of artificial intelligence analysis algorithms for medical images, more and more people have established a series of automatic diagnosis algorithms for DR. According to the purpose of detection, fundus image diagnosis algorithms can be divided into two categories.

One is to detect the fundus image based on the international diabetic retinopathy grade to determine the severity of the patient's DR. International diabetic retinopathy is divided into 5 grades: normal (no DR), mild non-proliferative DR (mild NPDR), moderate non-proliferative DR (moderate NPDR), severe non-proliferative DR (severe NPDR) and proliferative DR (PDR). In 2008, J. Nayak's team [6] used the traditional image feature extraction method to extract the features of the fundus images, and then inputted these features into the artificial neural network (ANN). They automatically classified 140 patients for their NPDR and PDR stages of DR. The final accuracy, sensitivity and specificity were 93%, 90% and 100% respectively. In 2016, Pratt's team [7] proposed a multi-layer convolutional neural network, which achieved 75% accuracy and 30% sensitivity on the dataset of diabetic retinopathy provided by Kaggle. In the same year, Google built a prediction model based on Inception-v3, and achieved an AUC of 0.991, a sensitivity of 90.3%, and a specificity of 98.1% on the EyePACS validation dataset. At the same time, the test classification of Google's model on the Messidor-2 dataset achieved an AUC of 0.990, a sensitivity of 0.870, and a specificity of 0.985 [8]. In 2018, Google improved its algorithm proposed in 2016. They increased the training dataset, enlarged the input size of the model, and improved the model architecture. The model predicted the 5-level grades of international diabetic retinopathy and achieved an AUC of 0.986, a sensitivity of 0.971, a specificity of 0.923 on the EyePACS validation dataset [9]. In 2019, Xu et al. [10] evaluated DR based on red lesions and bright lesions of fundus images. The model was tested with 19,904 fundus images. The AUC values of the model were PDR, 0.80; severe NPDR, 0.80; moderate NPDR, 0.77; and mild NPDR, 0.78.

The other is to detect the fundus images based on common lesions, and to use more accurate descriptions of the lesions in the fundus image. In 2010, García's team [15] used radial basis function (RBF) to detect red lesions in

the fundus images. The model they proposed was tested on 65 images and obtained an average sensitivity of 100%, an average specificity of 56.00% and an average accuracy of 83.08% on the image scale. In 2017, Tan's team [16] proposed a ten layers convolutional neural networks (CNN) for DR lesion detection on the pixel scale. The model achieved a sensitivity of 0.8758 and 0.7158 for exudates and dark lesions on 30,275,903 effective points of the CLEOPATRA database. It also achieved a sensitivity of 0.6257 and 0.4606 for hemorrhages and microaneurysms. In 2018, Khojasteh et al. [11] used a ten layers CNN to extract the deep features of the fundus image based on patch, and detected exudates, hemorrhages and microaneurysms, the results of the proposed approach shown overall accuracy for DIARETDB1 was 97.3% and 86.6% for e-Ophtha. In 2020, Pan's team [17] proposed a multi-label classification model for automatic analysis of fundus fluorescein angiography. The area under the curve (AUC) of the model reached 0.8703, 0.9435, 0.9647, and 0.9653 for detecting non-perfusion regions (NP), microaneurysms, leakages, and laser scars respectively on the dataset containing 4067 FFA images.

Common lesions related to DR in fundus images include microaneurysms, macular edema, hard lipid exudates, cotton wool-like soft exudates, etc. In real clinical diagnosis, these lesions may co-exist in fundus images. The simple solution to this multi-lesion detection is to treat each lesion independently, and to turn the multi-label classification problem into multiple binary classification problem. However, these methods ignore the potential relationship between lesions, so they are limited in nature.

In order to detect the multiple lesion labels in fundus images at the same time and to have a solid consideration about the complex topology of the lesion labels, we perform implicit modeling based on the correlation between the labels of graph convolutional network and build a fundus images multi-label classification model. First of all, we collect enough papers related to fundus lesions through keywords searching from China National Knowledge Infrastructure (CNKI) website [18] to construct a corpus. After that, we use the method of Global Vectors for Word Representation (Glove) to build a word vector model based on the corpus, and use the word vector model to construct a directed graph of the target label. Then we use the GCN network to model the label dependency. In the end, we use the convolutional neural network to extract image features and combined the output of the GCN network to simultaneously detect lesions. These lesions include laser scars, drusen, cup disc ratio ($C/D > 0.6$), hemorrhages, retinal arteriosclerosis, microaneurysms, hard exudates, soft exudates and so on.

Methods

Proposed model

We adjusted and improved ML-GCN proposed by chen’s team [19] to construct a multi-label classification model which would be suitable for fundus images. The model was composed of two parts: image feature extraction module and GCN-based classification module. The model frame is shown in Fig. 1, where $N = 8$ and $D = 2048$.

Image feature extraction module

The image feature extraction module used a CNN-based model to extract fundus image features. In our experiments, we tested different CNN models (VggNet [20], ResNet [21], DenseNet [22]), and finally decided to use ResNet-101 [21] to extract lesion features. Considering that some lesion features (microaneurysms, hard exudations, soft exudations, etc.) would be greatly difficult to recognize at low resolution, we used the 1024×1024 size fundus images as the input of ResNet-101. At this point we could get $2048 \times 32 \times 32$ feature maps from the “conv5_x” layer of ResNet-101, and then we used two convolution layers to downsample the feature maps, and the layers’ stride is 2, kernel size is 3×3 . Finally, we used the “adaptive max-pooling” to get the one-dimensional image features $F \in \mathbb{R}^D$, where $D = 2048$.

GCN-based classification module

We used the GCN-based classification module to build classifiers by modeling label dependencies. GCN network

is a kind of neural network that performs operations on graphs to learn a classification function [23, 24].

The GCN architecture used in this article is shown in Fig. 2. The input of GCN consists of two parts: Feature matrix $X \in \mathbb{R}^{N \times d}$ and Adjacency matrix $A \in \mathbb{R}^{N \times N}$. X is used to describe the characteristics of nodes, and A is a representative description of the graph structure, N is the number of categories, d is the number of features of the nodes.

Every hidden layer of GCN can be expressed as:

$$H^{l+1} = f(H^l, A) \tag{1}$$

where $H^l \in \mathbb{R}^{N \times d^l}$ is the graph-level outputs of the l th layer, d^l indicates the dimensionality of node features, H^0 is Feature matrix X , A is Adjacency matrix. Based on the propagation rule introduced in Kipf et al. [24], $f(\cdot)$ can be expressed as:

$$f(H^l, A) = \sigma(\hat{A}H^lW^l) \tag{2}$$

where $\hat{A} \in \mathbb{R}^{N \times N}$ is the normalized version of correlation matrix A , W^l is a weight matrix for the l th neural network layer, $\sigma(\cdot)$ is a non-linear activation function, In this experiment, we used LeakyReLU [25] as the activation function.

For the last layer, the output of GCN is $Z \in \mathbb{R}^{N \times D}$, $Z' \in \mathbb{R}^{D \times N}$ is obtained by Z transpose, D is the feature dimension of the final node, $D = 2048$. Finally, we apply the label features learned through GCN as a classifier to the image features, then we can obtain the predicted scores $y_{pred} \in \mathbb{R}^N$ as:

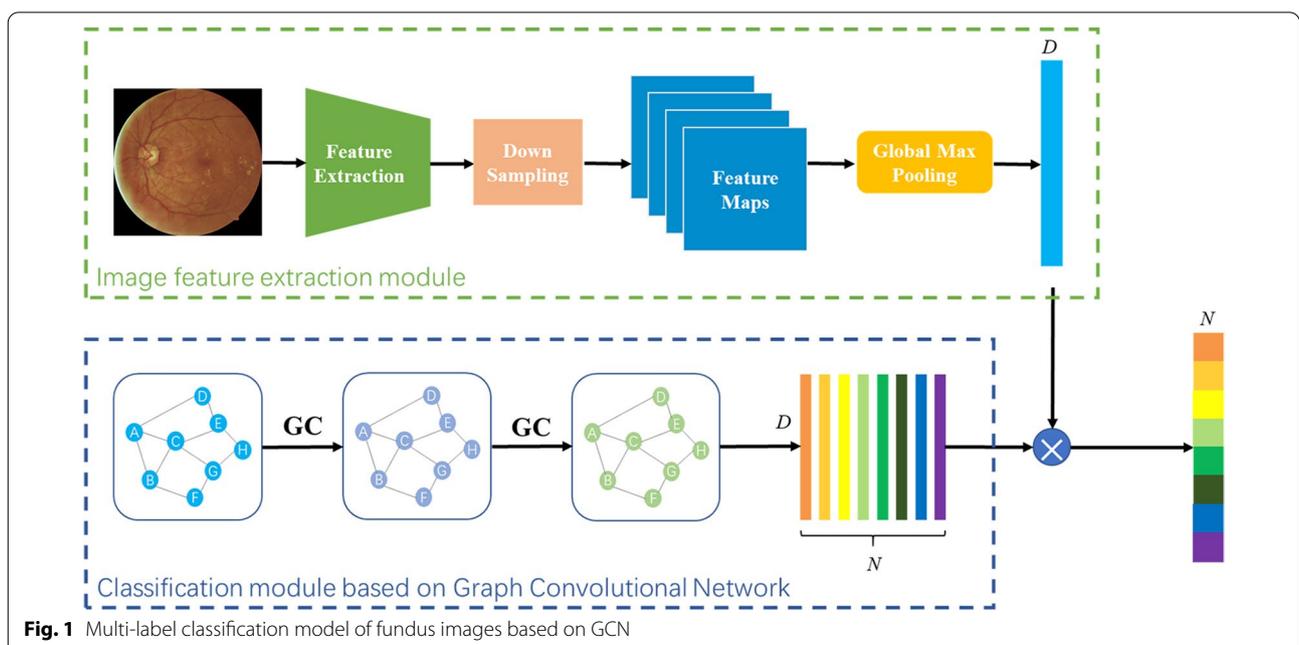


Fig. 1 Multi-label classification model of fundus images based on GCN

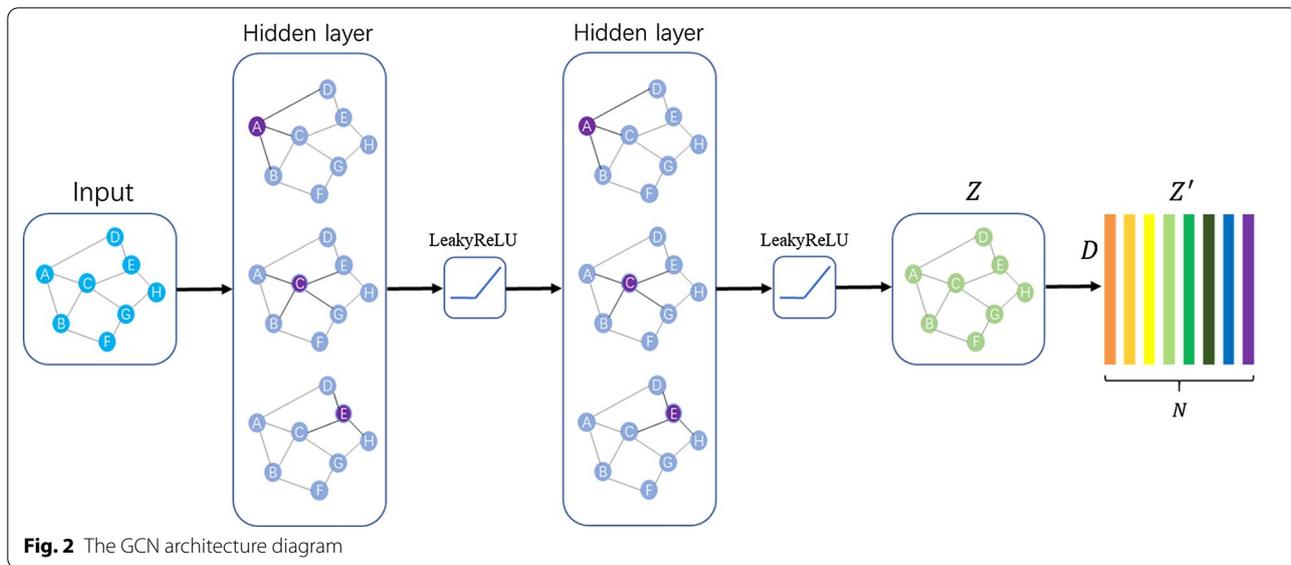


Fig. 2 The GCN architecture diagram

$$y_{pred} = FZ'. \tag{3}$$

The design of the GCN input matrix

In order to explore the complex topology between lesion labels, we used the GCN network to model label dependences. It can be seen from the above that the input of GCN is composed of feature matrix and adjacent matrix. We built the feature matrix based on the word embeddings of the labels, and built the adjacent matrix based on the co-occurrence pattern of the labels in the dataset. However, since it was difficult to obtain the word embeddings of the highly medical professional lesion labels based on a universal corpus, we created a professional fundus lesion-related corpus to obtain the word vectors of lesion labels.

The construction of feature matrix We collected articles related to fundus lesions through keywords searching from CNKI website and extracted the abstracts of these articles to construct a corpus related to fundus lesions. We have collected a total of 10,500 related documents. We then cleaned the corpus, including removing all symbols except commas and periods, replacing English abbreviations with full names, and so on. After that we segment the corpus based on the structured perceptron model [26] of HanLP [27], and the corpus after word segmentation contained Tokens 3M. In order to avoid the weight interference caused by stop words, we removed stop words from the corpus. After removing the stop words, the corpus contained Tokens 2M. In the end, we trained the Glove model [28] based on the processed corpus and generated the label word vector to construct the feature matrix $X \in \mathbb{R}^{N*d}$ of

the fundus lesion labels, where the word vector dimension d is set to 300.

The construction of adjacency matrix We referred to the method proposed by Chen et al. [19] to construct the adjacency matrix based on the conditional probability $P(L_i/L_j)$ between different labels, where L is the category label, and $P_{ij} = P(L_i/L_j)$ refers to the appearance probability of the label L_i when the label L_j appears in the Training set. Due to the use of the co-occurrence patternmode of labels in the training set to construct the adjacency matrix, the adjacency matrix will not be suitable for the test set. In order to improve the generalization ability of the model, we binarized P_{ij} to obtain the binarization adjacency matrix A' . At the same time, in order to avoid the over-smoothing of the label features caused by the binarization Adjacency matrix, we re-weighted the binarization adjacency matrix to obtain the final adjacency matrix. Therefore, the adjacency matrix can be expressed as:

$$A'_{ij} = \begin{cases} 0, & P_{ij} < p \\ 1, & P_{ij} \geq p \end{cases} \tag{4}$$

$$A_{ij} = \begin{cases} p / \sum_{j=1}^C A'_{ij}, & i \neq j \\ j \neq i \\ 1 - p, & i = j \end{cases} \tag{5}$$

where A' is the binarization adjacency matrix, A is the final adjacency matrix, and p is used to control the weight of the labels and its related labels. In this article, after experimental testing, we finally decided to set $\tau = 0.3, p = 0.25$.

Material and experiments

Dataset

The dataset we used came from the major special program for collaborative innovation in health care in Guangzhou. Data were obtained from Zhujiang Hospital of Southern Medical University, Third Affiliated Hospital of Sun Yat-sen University, Eastern Guangdong Hospitals and other Grade A class 3 hospitals between 2015 and 2018. The dataset consists of 2282 patients (1283 females, 999 males) and 7459 fundus image data (1887 cases of left eye, 1966 cases of right eyes). The image size is 1962×1944 . The patients are 20–93 years old, with an average age of $61.9147 (\pm 11.45)$. The age distribution is shown in Fig. 3.

The fundus image lesions to be labeled in this study include laser scars, drusen, cup disc ratio ($C/D > 0.6$), hemorrhages, retinal arteriosclerosis, microaneurysms, hard exudates and soft exudates. Each fundus image was annotated by two professional ophthalmologists. If there is no difference in the annotations, then the criteria are determined, otherwise they are discussed until consensus is reached. The distribution of fundus image labels is shown in Table 1.

Experimental design

Before training the model, we divided 7459 images into training set, validation set and test set according to a

At the same time, in order to verify the reliability of the created feature matrix, we built the feature vector of our model based on the other two pre-trained word vectors. One of the pre-trained word vectors is the 300d vectors which use the Glove model to train on Wikipedia-2014 and Gigaword5 datasets obtained by Pennington et al. [28]. Another pre-trained word vectors is 300d Chinese Word Vectors obtained by Li et al. [29] through using which used Skip-Gram model with Negative Sampling [30] trained on the Baidu Encyclopedia dataset. We constructed respectively the lesion label feature matrix $X^{en} \in \mathbb{R}^{N*d}$ and $X^{zh} \in \mathbb{R}^{N*d}$ by searching for the words contained in the English and Chinese lesion labels in the pre-trained word vectors. In addition, we also created a random matrix $X^r \in \mathbb{R}^{N*d}$ based on Gaussian distribution as a control to compare the performance of each feature matrix.

When training the model, we chose “stochastic gradient descent (SGD)” as the optimizer. Among the parameters of the optimizer, the momentum was 0.9, the weight attenuation coefficient was 0.0001, the learning rate for the pre-trained model ResNet-101 was 0.01, and the learning rate for other parts of the whole model was 0.1. At the same time, we chose “MultiLabelSoftMarginLoss” as the loss function of our model, which created a criterion that optimized a multi-label one-versus-all loss based on max-entropy. Its expression is as follows:

$$\text{loss}(y_{\text{pred}}, y) = -\frac{1}{N} * \sum_i y[i] * \log\left(\frac{1}{1 + e^{-y_{\text{pred}}[i]}}\right) + (1 - y[i]) * \log\left(\frac{e^{-y_{\text{pred}}[i]}}{1 + e^{-y_{\text{pred}}[i]}}\right) \quad (6)$$

proportion of 70%, 15%, and 15%. The number of each part and the label distribution information is shown in Table 2.

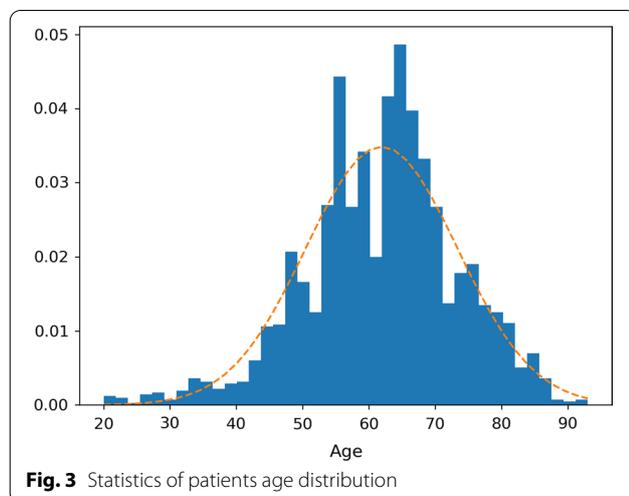


Fig. 3 Statistics of patients age distribution

where $y_{\text{pred}} \in \mathbb{R}^{B*N}$ is the prediction results of the model, B is the batch size, N is the number of label categories, $y[i] \in [0, 1]$ is the real label input by the model.

We used PyTorch (version = 1.4) to build the multi-label classification model of the fundus images proposed in this experiment. We used GPU graphics card (Nvidia GeForce GTX 1080Ti * 4) on Ubuntu16.04 system for model training, verification and testing.

Evaluation metrics

In order to evaluate the performance of the proposed model, Four metrics are used in this study: the average overall F1 Score (OF1), the average per-class F1 Score (CF1), the per-class accuracy (Acc) and the per-class area under the ROC curve (AUC). F1 Score is an index used in statistics to measure the accuracy of classification models, which takes into account both model accuracy and recall rate. F1 Score can be regarded as an average weighting of model accuracy and recall. The value of F1 ranges from 0 to 1 and the higher the value, the better the performance of the model. AUC is an important curve to

measure the classification problem. In this study, AUC was calculated to judge the performance of the model's ability to classify each class of the lesion label. The closer the value of AUC is to 1, the better performance the model of each lesion label's classification has.

For each image, we assigned labels with confidence greater than 0.5 to be positive, and compared them with a ground-based true-value labels. These measures do not need a fixed number of labels per image.

The calculation formula for each indicator is as follows:

$$OP = \frac{\sum_i TP^i}{\sum_i (TP^i + FP^i)} \quad (7)$$

$$OR = \frac{\sum_i TP^i}{\sum_i (TP^i + FN^i)} \quad (8)$$

$$OF1 = 2 * \frac{OP * OR}{OP + OR} \quad (9)$$

$$CP = \sum_i \frac{TP^i}{TP^i + FP^i} / N \quad (10)$$

$$CR = \sum_i \frac{TP^i}{TP^i + FN^i} / N \quad (11)$$

$$CF1 = 2 * \frac{CP * CR}{CP + CR} \quad (12)$$

$$Acc[i] = \frac{TP^i + TN^i}{TP^i + FN^i + TN^i + FP^i} \quad (13)$$

where $i \in [1, \dots, N]$, N is the number of lesion labels, the true positive i (TP^i) indicates that the number of fundus images where the i_{th} class of lesion labels should have existed in the image and was predicted by the model to exist, the false negative i (FN^i) indicates that the number of fundus images where the i_{th} class of lesion labels should have existed in the image but was predicted by the model to not exist, the true negative i (TN^i) indicates that the number of fundus images where the i_{th} class of lesion labels should not have existed in the image and was predicted by the model to not exist, the false positive i (FP^i) indicates that the number of fundus images where the i_{th} class of lesion labels should not have existed in the image but was predicted by the model to exist, OP and OR are the average overall precision and recall, CP and CR are the average per-class precision and recall.

Results and discussion

The performance results of the model training based on different feature matrices are shown in Tables 3 and 4. X is the feature matrix constructed in Section "The Construction of Feature Matrix" based on the fundus lesion related corpus. X^{rd} , X^{en} , and X^{ch} are respectively the feature matrix constructed in Section "Experimental Design" based on the random distributions, Wikipedia 2014, and Baidu Encyclopedia.

Among them, $OF1$ and $CF1$ values of model based on X^{rd} as the baseline are 0.483 and 0.389 respectively with the worst performance. $OF1$ and $OF1$ values of model based on X^{en} are 0.627 and 0.570 respectively. $OF1$ and $OF1$ values of model based on X^{ch} are 0.633 and 0.582 respectively. The performance of model based on X^{ch} is not much different from that of model based on X^{en} . $OF1$ and $OF1$ values of model based on X are 0.808 and 0.792 respectively with the best performance. By comparing $OF1$ and $OF1$ values of different models, it can be seen that the word vector model based on a universal corpus cannot describe the labels of fundus lesions. The feature matrix constructed based on the fundus lesions related corpus can more comprehensively and accurately represent the complex co-occurrence relationship between lesion labels, which makes model based on X to have better performance. Figure 4 shows the AUC curve of each label of the model based on X .

The Acc and ROC values of models shown that the model had a better detection results for Laser scars, Drusen and Haemorrhages, but had a poor detection ability for microaneurysms, soft exudates and hard exudates. We speculated that the reason is that microaneurysms looked just like small, red spots in the retinal capillaries [31]. It was difficult for the model to distinguish microaneurysms from the background of the fundus images, especially when the original image was reduced when input. The 2017 report of Tan et al. [16] also verified our opinions. Tan et al. proposed a ten layers CNN architecture for DR lesion segmentation, achieved a sensitivity of 0.46 for segmentation of microaneurysms. Tan's job shown that microaneurysms were very difficult to distinguish from the surrounding background pixels. For soft exudates and hard exudates, we found that these two kinds of lesions often accompanied multiple other fundus lesions at the same time, which made the model difficult to extract the features of all fundus lesions. In 2018, lam et al. [32] used image patches to detect retinal lesion which can avoid the interference of different fundus lesions, the proposed model achieved the AUC of 0.94 and 0.95 for detection of microaneurysms and exudates. At the same time, in order to verify the superiority of the model, we also made a comparison with the

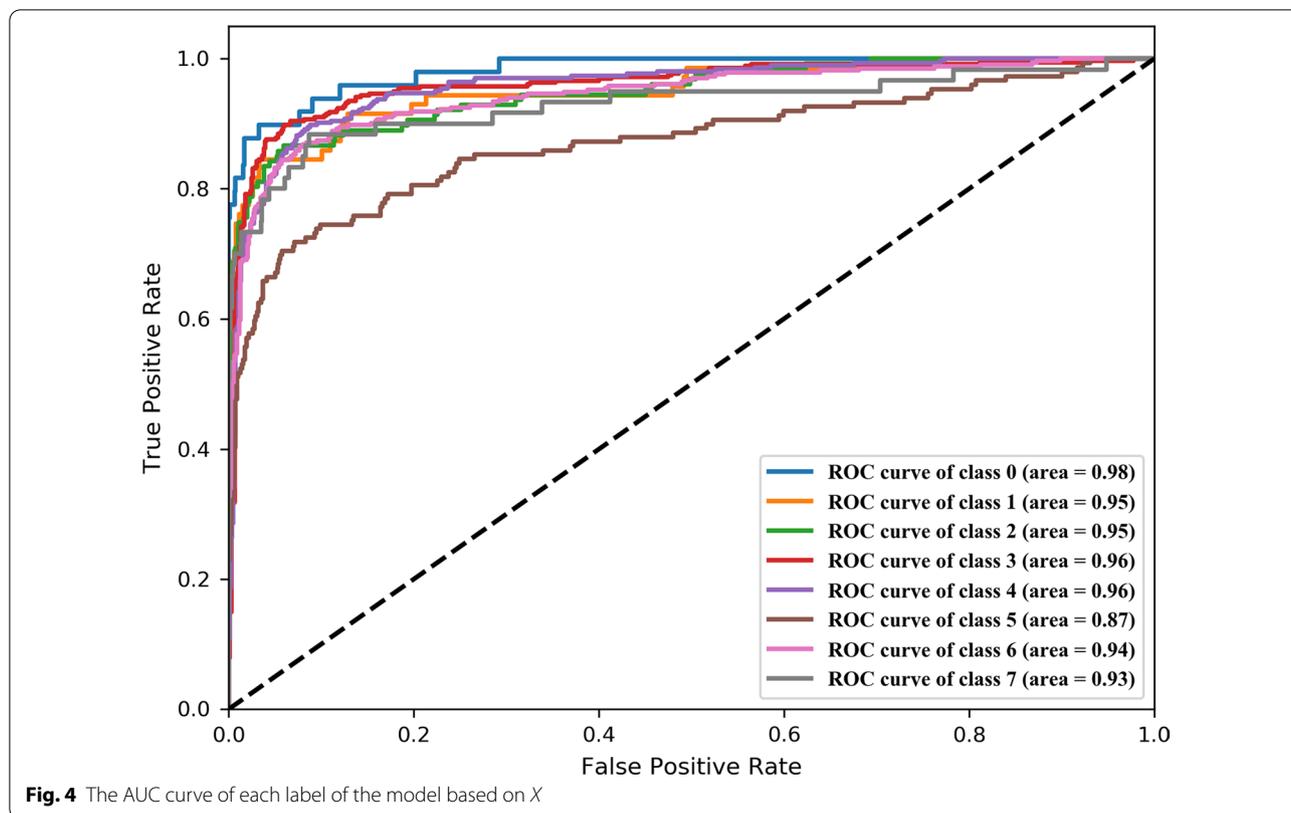


Table 1 Statistics of image label distribution

Label	Age	Age standard deviation	Sex ratio (M/T)	Number
Laser scars	62.73	12.62	0.37	232
Drusen	62.63	10.66	0.35	590
Cup disc ratio	57.09	15.45	0.51	632
Hemorrhages	59.52	11.48	0.44	2149
Retinal arteriosclerosis	64.48	10.95	0.43	2686
Microaneurysms	57.47	9.91	0.49	997
Hard exudates	60.91	10.74	0.41	1857
Soft exudates	59.09	10.38	0.44	467

Table 2 Fundus image label distribution

	Training	Validation	Test	Total
Laser scars	139	49	44	232
Drusen	422	71	97	590
Cup disc ratio	394	127	111	632
Haemorrhages	1400	359	390	2149
Retinal arteriosclerosis	2073	304	309	2686
Microaneurysms	705	151	141	997
Hard exudates	1201	335	321	1857
Soft exudates	324	61	82	467

multi-label classification model of the fundus lesions proposed by Pan et al. [17]. Pan trained a multi-label classification DenseNet model based on FFA images to detect NP, microaneurysms, leakages, and laser scars with AUC of 0.880, 0.980, 0.974 and 0.967. We used fundus photographs of the retina to build a multi-label classification model. Compared to FFA images, the fundus photographs of the retina were more often used for large-scale fundus disease screening, but were more difficult for accurate fundus disease detection.

Table 3 OF1 and CF1 values of the models based on different feature matrices

	Model base on X	Model base on X^{rd}	Model base on X^{en}	Model base on X^{ch}
OF1	0.808	0.483	0.627	0.633
CF1	0.792	0.389	0.570	0.582

Table 4 Acc and ROC values of the models trained based on different feature matrices

	Model base on X		Model base on X^{rd}		Model base on X^{en}		Model base on X^{ch}	
	Acc	ROC	Acc	ROC	Acc	ROC	Acc	ROC
Laser scars	0.986	0.982	0.954	0.911	0.972	0.914	0.974	0.963
Drusen	0.971	0.952	0.945	0.807	0.949	0.880	0.948	0.872
Cup disc ratio	0.956	0.949	0.902	0.829	0.921	0.879	0.919	0.840
Haemorrhages	0.926	0.961	0.768	0.776	0.836	0.879	0.829	0.871
Retinal arteriosclerosis	0.914	0.957	0.719	0.742	0.783	0.813	0.805	0.837
Microaneurysms	0.921	0.871	0.847	0.698	0.885	0.823	0.897	0.831
Hard exudates	0.903	0.944	0.805	0.804	0.847	0.889	0.845	0.870
Soft exudates	0.974	0.930	0.935	0.730	0.955	0.845	0.957	0.808

Our multi-labels classification model based on GCN can detect eight types of fundus lesions at the same time, and the AUC of detecting Laser scars was higher than the model proposed by Pan et al.

Conclusions

In the actual clinical diagnosis, there would be a variety of lesions in the fundus images. Therefore, different from the existing single classification model, we proposed a multi-label classification model based on GCN. We also tested a variety of fundus photographs of the retina to make it better applied in the real medical scene. Our experimental results on the constructed multi-center clinical data set demonstrate the promising performance and broad application of the proposed model.

In summary, we proposed a multi-label classification model based on GCN, which enable the model to learn the complex topology between lesion labels. At the same time, we constructed a related to fundus lesions corpus and verified its superiority through comparison. The multi-label classification model had been trained and verified on the real datasets. It can detect 8 types of lesions such as laser scars, drusen, cup disc ratio ($C/D > 0.6$), hemorrhages, retinal arteriosclerosis, microaneurysms, hard exudates and soft exudates in fundus images very well. Therefore, our multi-label classification model can well assist ophthalmologists in the diagnosis of DR, reduce the workload of ophthalmologists in clinical practice, and improve the diagnostic efficiency of ophthalmologists.

Abbreviations

Acc: Accuracy; ANN: Artificial neural network; AUC: Area under the ROC curve; CF1: Average per-class F1 score; CNKI: China national knowledge infrastructure; CNN: Convolutional neural network; CP: Average per-class precision; CR: Average per-class recall; C/D: Cup disc ratio; DR: Diabetic retinopathy; FFA: Fluorescein fundus angiography; GCN: Graph convolutional network; Glove: Global Vectors for Word Representation; OF1: Average overall F1 score; OP: Average overall precision; OR: Average overall recall; NP: Non-perfusion; NPDR: Non-proliferative diabetic retinopathy; PDR: Proliferative diabetic retinopathy; RBF: Radial basis function; SGD: Stochastic gradient descent.

Acknowledgements

The authors would like to thank Liangjun Zhang for insightful discussion and language assistance. And the authors thank the anonymous reviewers for their careful review and valuable comments.

About this supplement

This article has been published as part of BMC Medical Informatics and Decision Making Volume 21, Supplement 2 2021: Health Big Data and Artificial Intelligence. The full contents of the supplement are available at <https://bmcmmedinformdecismak.biomedcentral.com/articles/supplements/volume-21-supplement-2>.

Authors' contributions

Study concept and data provision: YZ, YLC and MNM. Data preparation and processing: MNM and XYL. Formulation of model: YLC and YZ. Wrote and revised the manuscript: YLC and MNM. The manuscript review: YLC, XYL and YZ. All the authors read and approved the final version of the manuscript.

Funding

This research/work was supported by the National Key Research and Development Program of China (Nos. 2018YFC0116900, 2016YFC0901602), Key Research and Development Program of Guangdong (No. 2018B010109006), National Natural Science Foundation of China (NSFC) (No. 61876194), Science and Technology Program of Guangzhou (No. 202011020004), 111 Project (Grant No.B12003).

Availability of data and materials

The data is collected and integrated by the company based on the Science and Technology Program of Guangzhou (No. 201604020016). Because of the company's regulations, the data cannot be made publicly available.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing financial interests.

Author details

¹ School of Biomedical Engineering, Sun Yat-sen University, No. 132 Waihuan East Road, Guangzhou 510006, China. ² Department of Medical Informatics, Zhongshan School of Medicine, Sun Yat-sen University, No. 74 Zhongshan 2nd Road, Guangzhou 510080, China. ³ Zhongshan School of Medicine, Sun Yat-sen University, No.74 Zhongshan 2nd Road, Guangzhou 510080, China. ⁴ Key Laboratory of Tropical Disease Control (Sun Yat-sen University), Ministry of Education, No. 74 Zhongshan 2nd Road, Guangzhou 510080, China.

Received: 14 January 2021 Accepted: 8 February 2021

Published: 30 July 2021

References

- Ogurtsova K, da Rocha Fernandes J, Huang Y, Linnenkamp U, Guariguata L, Cho NH, Cavan D, Shaw J, Makaroff L. Idf diabetes atlas: global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract.* 2017;128:40–50.
- Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, Chen S-J, Dekker JM, Fletcher A, Grauslund J, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care.* 2012;35(3):556–64.
- Group ETDRSR, et al. Early photocoagulation for diabetic retinopathy: Etdrs report number 9. *Ophthalmology.* 1991;98(5):766–85.
- Kornblau IS, El-Annan JF. Adverse reactions to fluorescein angiography: a comprehensive review of the literature. *Surv Ophthalmol.* 2019;64(5):679–93.
- Kwiterovich KA, Maguire MG, Murphy RP, Schachat AP, Bressler NM, Bressler SB, Fine SL. Frequency of adverse systemic reactions after fluorescein angiography: results of a prospective study. *Ophthalmology.* 1991;98(7):1139–42.
- Nayak J, Bhat PS, Acharya R, Lim CM, Kagathi M. Automated identification of diabetic retinopathy stages using digital fundus images. *J Med Syst.* 2008;32(2):107–15.
- Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. *Proc Comput Sci.* 2016;90:200–5.
- Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316(22):2402–10.
- Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, Peng L, Webster DR. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology.* 2018;125(8):1264–72.
- Xu Y, Wang Y, Liu B, Tang L, Lv L, Ke X, Ling S, Lu L, Zou H. The diagnostic accuracy of an intelligent and automated fundus disease image assessment system with lesion quantitative function (smarteeye) in diabetic patients. *BMC Ophthalmol.* 2019;19(1):184.
- Khojasteh P, Aliahmad B, Kumar DK. Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. *BMC Ophthalmol.* 2018;18(1):1–13.
- Alban M, Gilligan T. Automated detection of diabetic retinopathy using fluorescein angiography photographs. Stanford: Report of Stanford Education; 2016.
- Haloj M. Improved microaneurysm detection using deep neural networks. Preprint; 2015. [arXiv:1505.04424](https://arxiv.org/abs/1505.04424).
- Sinthanayothin C, Kongbunkiat V, Phoojaruenchanachai S, Singalavanija A. Automated screening system for diabetic retinopathy. In: 3rd international symposium on image and signal processing and analysis, 2003. ISPA 2003. proceedings of the, 2003; vol. 2. New York: IEEE. p. 915–20.
- García M, López MI, Álvarez D, Hornero R. Assessment of four neural network based classifiers to automatically detect red lesions in retinal images. *Med Eng Phys.* 2010;32(10):1085–93.
- Tan JH, Fujita H, Sivaprasad S, Bhandary SV, Rao AK, Chua KC, Acharya UR. Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. *Inf Sci.* 2017;420:66–76.
- Pan X, Jin K, Cao J, Liu Z, Wu J, You K, Lu Y, Xu Y, Su Z, Jiang J, et al. Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning. *Graefes Arch Clin Exp Ophthalmol.* 2020;258(4):779–85.
- China National Knowledge Infrastructure. <https://www.cnki.net>. Accessed 20 Nov 2020.
- Chen Z-M, Wei X-S, Wang P, Guo Y. Multi-label image recognition with graph convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2019. p. 5177–86.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556); 2014.
- 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; 2016. p. 770–8.
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–8.
- Duvenaud DK, Maclaurin D, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, Adams RP. Convolutional networks on graphs for learning molecular fingerprints. In: Advances in neural information processing systems; 2015. p. 2224–32.
- Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. Preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907); 2016.
- Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In: Proc. Icm1, vol. 30; 2013. p. 3.
- Collins M. Discriminative training methods for hidden Markov models: theory and experiments with perceptron algorithms. In: Proceedings of the 2002 conference on empirical methods in natural language processing (EMNLP 2002); 2002. p. 1–8.
- Han Language Processing. <https://github.com/hankcs/HanLP>. Accessed 20 Nov 2020.
- Pennington J, Socher R, Manning CD. Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP); 2014. p. 1532–43.
- Li S, Zhao Z, Hu R, Li W, Liu T, Du X. Analogical reasoning on Chinese morphological and semantic relations. Preprint [arXiv:1805.06504](https://arxiv.org/abs/1805.06504); 2018.
- Goldberg Y, Levy O. Word2vec explained: deriving Mikolov et al's negative-sampling word-embedding method. Preprint [arXiv:1402.3722](https://arxiv.org/abs/1402.3722); 2014.
- Manjaramkar A, Kokare M. Statistical geometrical features for microaneurysm detection. *J Digit Imaging.* 2018;31(2):224–34.
- Lam C, Yu C, Huang L, Rubin D. Retinal lesion detection with deep learning using image patches. *Invest Ophthalmol Vis Sci.* 2018;59(1):590–6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

