

Improved Model of Eye Disease Recognition Based on VGG Model

Ye Mu^{1,2,3,4}, Yuheng Sun¹, Tianli Hu^{1,2,3,4}, He Gong^{1,2,3,4}, Shijun Li^{1,2,3,4,*} and Thobela Louis Tyasi⁵

¹College of Information Technology, Jilin Agricultural University, Changchun, 130118, China

²Jilin Province Agricultural Internet of Things Technology Collaborative Innovation Center, Changchun, 130118, China

³Jilin Province Intelligent Environmental Engineering Research Center, Changchun, 130118, China

⁴Jilin Province colleges and universities The 13th Five-Year Engineering Research Center, Changchun, 130118, China

⁵Department of Agricultural Economics and Animal Production, University of Limpopo, 0727, Polokwane, South Africa

*Corresponding Author: Shijun Li. Email: lishijun@jlau.edu.cn

Received: 05 January 2021; Accepted: 14 February 2021

Abstract: The rapid development of computer vision technology and digital images has increased the potential for using image recognition for eye disease diagnosis. Many early screening and diagnosis methods for ocular diseases based on retinal images of the fundus have been proposed recently, but their accuracy is low. Therefore, it is important to develop and evaluate an improved VGG model for the recognition and classification of retinal fundus images. In response to these challenges, to solve the problem of accuracy and reliability of clinical algorithms in medical imaging this paper proposes an improved model for early recognition of ophthalmopathy in retinal fundus images based on the VGG training network of densely connected layers. To determine whether the accuracy and reliability of the proposed model were greater than those of previous models, our model was compared to ResNet, AlexNet, and VGG by testing them on a retinal fundus image dataset of eye diseases. The proposed model can ultimately help accelerate the diagnosis and referral of these early eye diseases, thereby facilitating early treatment and improved clinical outcomes.

Keywords: Deep learning model; dense block; eye disease recognition; fundus retina image; VGG

1 Introduction

Using computers for disease diagnosis has become a focus of medical research. In many instances, characteristics of diseases are determined based on an expert's judgment of various disease images, and measures such as referral treatment are taken after early diagnosis to prevent the development and spread of the disease. Therefore, to help reduce the expense of this procedure and increase its speed and accuracy it is important to develop a high-precision eye disease automatic detection and recognition model that incorporates deep learning. Structural feature-based methods define high-quality retinal images as retinal images with separable anatomical structures [1]. Examples include the retinal vasculature and the optic disc. Those methods are closely related to ophthalmic diagnosis, but they require accurate



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

anatomical segmentation. However, complex segmentation is complex and error-prone, especially for poor-quality images [2].

Glaucoma is an eye disease that can cause nerve damage and irreversible vision loss as it develops, so early screening and diagnosis of glaucoma are very important to maintain vision [3]. As a population ages, the number of patients with glaucoma increases every year. Although many studies now use deep learning to diagnose the early segmentation of glaucoma and have improved the diagnosis of early glaucoma [4], the recognition of eye disease still relies on the expertise of professional ophthalmologists for segmentation [5]: if nonexperts are doing the diagnoses, an important diagnosis of eye disease patients may be missed. Fortunately, machine learning has achieved significant success in eye disease recognition. These achievements include, but are not limited to, early diagnosis of glaucoma [6], but also diagnosis of myopia and cataracts, and detection and analyses of retinal fundus images and optical coherence tomography images [7]. Therefore, a highly accurate automatic diagnosis model for eye disease based on deep learning algorithms is required to overcome the shortcomings of traditional methods.

2 Related Work

The detection of eye disease is divided into features based on manual segmentation of the vertical cup-to-disc ratio [8] and recognition based on a convolutional neural network (CNN) [9]. Full [10] proposed a new optic disc perception integrated network based on global fundus images and local optic disc regions. Giancardo et al. [11] proposed a new set of features independent of visual field or resolution to describe the morphology of a patient's blood vessels. Preliminary results show that those features can be used to estimate image quality in an order of magnitude faster than previous techniques. Juneja et al. [12] proposed an artificial eye disease expert system based on optic disc and cup segmentation. A CNN-based deep learning architecture was developed for the automatic detection of eye disease. Fan et al. [13] proposed a new normalization technique to extract four new features related to blood vessel brightness reflection, and they verified the features using the accuracy of a linear discriminant analysis classifier. Fan et al. [14] extracted many features of the central point of the blood vessels vessel and applied a feature selection technology based on genetic search to obtain the optimal feature subset of an artery/vein classification.

In past developments, deep CNNs gained considerable attention in many image processing fields, including retinal image quality assessment [15,16] and automatic screening of retinopathy [17,18]. Yu et al. [19] proposed a new retinal image quality classification (IQC) method that combined unsupervised features of saliency mapping with supervised features of CNNs, and put them into a support vector machine: high and low-quality retinal fundus pictures were automatically detected. He et al. [20] proposed a residual learning framework to simplify the training of the network, which was much deeper than a previously used network. The distributed learning architecture proposed by Chen et al. [21] included six learning layers: four convolutional layers and two fully connected layers, which could be used to further increase the accuracy of the diagnosis of glaucoma. Li et al. [22] proposed a new classification-based eye disease detection method that used a deep CNN generated from a large-scale general dataset to represent visual appearance, and combined global and local features to reduce the effects of mismatch. Zhou et al. [23] proposed a road traffic sign recognition algorithm when signs were obscured by poor illumination, occlusion, loss and deformation, and proposed an improved VGG (IVGG) model based on the VGG model. Su et al. [24,25] proposed a tag identification protocol based on ALOHA, which improved the reading efficiency of a UHF RFID system based on the EPC C1 Gen2 air interface protocol, and proposed a rapidly evolving network policy incremental algorithm, which overcame the limitations of traditional algorithms software-defined network architecture. Sudha et al. [26] proposed a deep CNN that could detect defects such as retinal exudates, and then classify and identify conditions. The structure tensor could be effectively used to enhance the local pattern of edge elements

and change the intensity of the edge of the object. The gradient descent method was used for active contour approximation from image segmentation. Fauw et al. [27] proposed a novel deep learning architecture for a set of clinically heterogeneous 3D optical coherence tomography images, and Cui et al. [28] proposed a CNN based on deep photography images automatic recognition and effective computer methods. Quan et al. [29] distinguished natural images and computer-generated images and proposed an effective method based on CNNs, and put forward a can adapt to different size of the input image patches at the same time keep the depth of the fixed network, the network is introduced into the so-called local to the global strategy.

Therefore, the recognition task of eye disease recognition and classification should essentially be the classification of high-frequency image features, where generated forensic images and the domain knowledge of natural images and the generated image forensics can be used. When dense link blocks in DenestNet are used to improve the VGG network training model to experiment on an eye disease dataset, the combination of retinal fundus retina recognition and eye disease assessment can be used to more accurately identify early eye disease from retinal images.

3 Experiments and Discussions

3.1 Datasets Used

This experiment used the glaucoma dataset iChallenge-GON, the cataract dataset iChallenge-AMD, and the myopia dataset iChallenge-PM provided by the Baidu Research Open Access dataset. All three datasets were collected using two different fundus cameras: for training, a Zeiss VISUCAM 500 (2124 × 2056 pixels), and for verification and testing, a Canon CR-2 (1634 × 1634 pixels). All images of the three datasets had both a macula and an optic disc. To address this challenge, three eye disease datasets were used, each of which had 1,200 color fundus photographs to choose from, including fundus photographs from patients with no eye diseases (90%) and with eye diseases (10%). All fundus images were stored as JPEG files. The dataset was divided into three subsets in accordance with state the ratios, which were used for training, verification, and testing respectively. The incidence of eye diseases was the same after the stratification. Part of the data and functions required for the eye disease identification experiment are shown in Fig. 1.

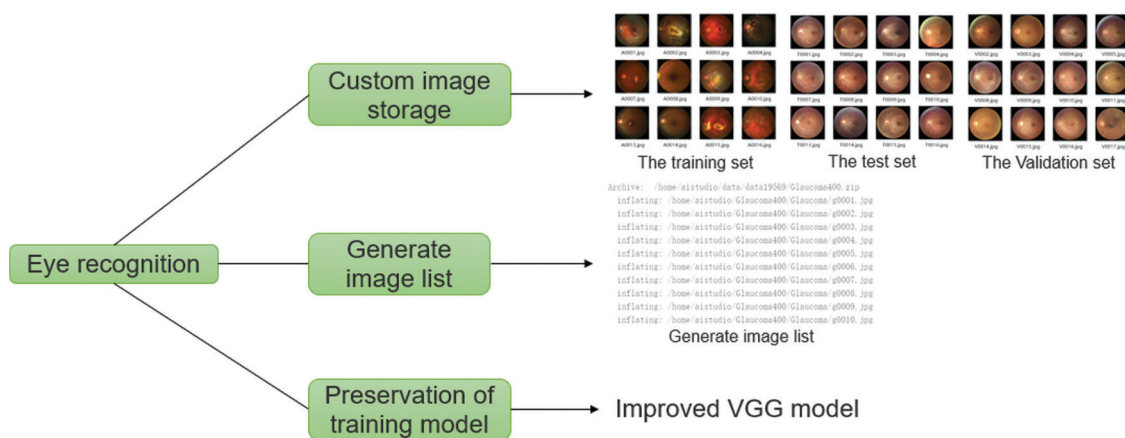


Figure 1: Part of the data and functions required for the eye disease identification experiment

3.2 Improved VGG Model

In the deep learning network, the VGG network model used a smaller convolution kernel based on AlexNet and deepened the network to achieve better results. The structure of the VGG network training

model is shown in Fig. 2. However, with the deepening of the network, the problem of gradient disappearance became more and more obvious. Many papers have proposed methods to solve this problem. For example, ResNet uses identity mapping in every two convolutional layers to reduce the disappearance of gradients caused by deep networks. By making extensive use of functions, DenseNet can obtain better results and fewer parameters and directly connect all layers under the premise of ensuring the maximum transmission of information between the various layers in the network. One of the advantages of DenseNet is that the network is narrower and has fewer parameters. Therefore, the effect of its use of dense connection blocks to improve the VGG network model is better than that of ResNet, AlexNet, and VGG. The proposed basic idea of improving the VGG model with dense blocks is the same as that of ResNet, but it establishes a close connection between the front and rear layers. Also, relative to

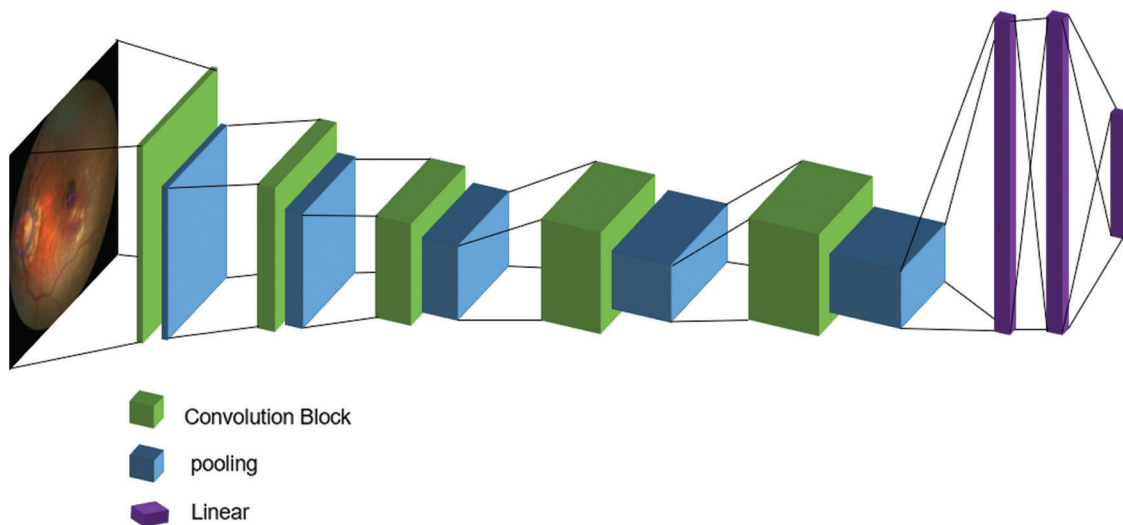


Figure 2: VGG network training model

ResNet, it achieves better performance with fewer parameters and lower computational cost.

Adding the dense blocks in DenseNet to the original VGG training network creates a training network that can effectively distinguish different categories. The mentioned gradient disappearance problem is that the deeper the network depth, the more easily the gradient disappears. This is because the input information and gradient information are transmitted between layers. The effect of the dense block of each layer is that it is equivalent to directly connecting the input and loss, so the gradient can be reduced to build a deeper network disappearance, enhance the spread of functions, promote the reuse of features, and greatly reduce the number of parameters. A VGG network training model improved by dense link blocks connects all layers in a narrow connection mode to ensure the maximum information flow between the layers. After input data preprocessing steps such as image scale changes and RGB changes, the densely connected layer is used as a series of feature extractors for various functions of computer vision in machine learning. The dense blocks make the transmission of features and gradients more effective, and the network is easier to train. Each layer can directly use the gradient of the loss function and its initial input. This is implicit deep supervision, which helps to train deeper networks.

Unlike ResNet, the proposed VGG network training model based on dense blocks would not be merged but would be connected before being passed to the layers. Therefore, the L layer has L inputs, which are the feature maps of all convolutional blocks before the l layer, and its own feature maps are passed to all subsequent layers.

Let X_l be the output of the l th layer. The traditional VGG connection method may have l layers, and there are corresponding l direct connections between each layer and its subsequent layers, while a dense block has $l(l + 1)/2$ direct connections. The input of the first layer is X_0 , of the second layer is X_0 and X_1 , and so on. X_{l1} and X_{l2} are the outputs of VGG and the dense blocks respectively. The network structure of the proposed model is shown in Fig. 3.

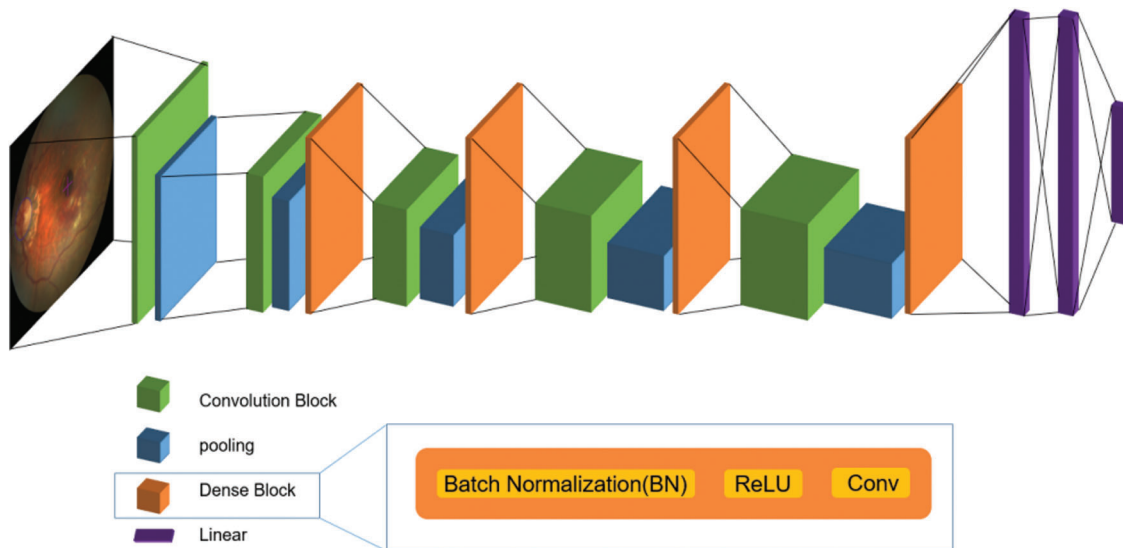


Figure 3: The network structure diagram of the improved VGG network training model

$$X_{l1} = H_{l1} * (X_l - 1) \tag{1}$$

$$X_{l2} = H_{l2} * X_0, X_1, \dots, X_l - 1 \tag{2}$$

Because the DenseNet dense blocks do not need to relearn redundant feature maps and have fewer feature maps, the number of parameters in the network is smaller and the calculation is faster. The traditional feedforward structure can be regarded as an algorithm for state transfer between layers. Each layer receives the state of the previous layer, and then passes the new state to the next layer, conveying the information that must be retained. In the structure of a dense block, the information added to the network is significantly different from the reserved information. Each convolutional layer in the densely connected blocks is very narrow, only a few feature maps are added to the network, and the feature maps remain unchanged. Finally, the classifier makes predictions based on all feature maps in the network.

When it comes to dense blocks, H_{l2} represents a series of composite functions: batch normalization (BN), rectified linear unit (ReLU), and convolution (Conv), which are classified and simplified to BN-ReLU-Conv. The concatenation of these composite functions is represented as a dense block, and then a new feature map is created. Therefore, the feature maps of all previous layers can be used by subsequent layers, which improves the information flow and the gradient of the entire network.

The network used was a randomly initialized binary classifier training network. Then, the preprocessed retinal image was used to fine-tune the model. The improved VGG training network based on dense blocks used in the experiment had three dense blocks, and all blocks had the same number of layers. Before entering the first dense block, the input image first passed through a 3×3 convolutional layer. After the transition layer between the two dense blocks, a 1×1 convolutional layer and a 2×2 average pooling layer were added. After the last dense block, a global average pooling layer was used. Through random

initialization, the result of fine-tuning was much better than that of the architecture. The binary classifier network configuration is shown in [Tab. 1](#).

Table 1: Classifier network configuration in improved model

Configuration	Classification
Convolution	3×3 Conv1 (stride 1)
Pooling	3×3 (max pooling) (stride 2)
Dense block 1	3×3 Conv (stride 1)
Transition layer 1	1×1 Conv 2×2 (average pooling) (stride 2)
Dense block 2	$3 \times (3 \times 3)$ Conv (stride 1)
Transition layer 2	1×1 Conv 2×2 (average pooling) (stride 2)
Dense block 3	$3 \times (3 \times 3)$ Conv (stride 1)
Classification layer	7×7 (average pooling) fully connected

¹Each Conv shown in the table corresponds to BN-ReLU-Conv. All networks and parameters were updated and optimized by stochastic gradient descent.

3.3 Experimental Results

The experiment implementation process comprised data preparation, network configuration, network model training, model evaluation, and model prediction. The improved model was trained on the iChallenge-GON dataset, and the training of the ResNet, AlexNet, and VGG network models on the three datasets were compared to evaluate the proposed model.

[Tabs. 2 to 4](#) show that the proposed VGG model, relative to ResNet, AlexNet, and VGG models, was more accurate with fewer errors.

Table 2: Proposed model accuracy and error comparison as trained on the iChallenge-GON dataset

Methods	ResNet	AlexNet	VGG	Proposed model
Accuracy (%)	0.8714998	0.9269998	0.937500	0.940624
Cost (%)	0.3055544	0.2712584	0.216918	0.222547

Table 3: Proposed model accuracy and error comparison as trained on iChallenge-AMD dataset

Methods	ResNet	AlexNet	VGG	Proposed model
Accuracy (%)	0.75000	0.87500	0.90625	0.968750
Cost (%)	0.33185	0.30455	0.27440	0.132812

By observing the errors and accuracy of the output during the training process of the proposed VGG network training model, the results of the network training can be evaluated. [Fig. 4](#) shows that the average errors in the training process gradually declined, while the training accuracy gradually improved.

Table 4: Proposed model accuracy and error comparison as trained on iChallenge-PM dataset

Methods	ResNet	AlexNet	VGG	Proposed model
Accuracy (%)	0.719998	0.926998	0.877600	0.942967
Cost (%)	0.534796	0.271258	0.227960	0.188623

**Figure 4:** Accuracy and error rate curves of the proposed model in the three datasets

4 Conclusion

The VGG network model uses a smaller convolution kernel based on AlexNet and deepens the network to achieve better results. ResNet uses identity mapping in every two convolutional layers to reduce the disappearance of gradients caused by deep networks. DenseNet obtains better results and fewer parameters through extensive use of functions. The central idea is to directly connect all layers to ensure maximum information transmission between all layers in the network. One of the advantages of dense blocks is that the network is narrower and has fewer parameters. Therefore, in this paper an improved VGG eye disease classification and recognition model based on dense blocks is proposed. To effectively distinguish different categories, the proposed model uses various features of computer vision and a series of feature extractions to enhance machine learning. The proposed network can reduce the gradient disappearance problem, enhance the spread of functions, promote the reuse of functions, and greatly reduce the number of parameters. The effect of using dense blocks to improve the VGG network model is better than that of ResNet, AlexNet, and VGG. The proposed model was compared with those networks and evaluated on the same datasets. In the iChallenge-GON dataset, the overall accuracy of the proposed model's classification and recognition was 0.94, which was higher than ResNet's 0.87, AlexNet's 0.92, and VGG's 0.93. In the iChallenge-AMD dataset, the overall accuracy of the proposed model was 0.96, surpassing ResNet's 0.75, AlexNet's 0.87, and VGG's 0.90. In the iChallenge-PM dataset, the overall accuracy was 0.94, which exceeded ResNet's 0.71, AlexNet's 0.92, and VGG's 0.87. Experimental results show that the improved model can accurately distinguish retinal images of eye diseases and was better than other popular deep learning networks. Also, in ophthalmology, the improved model can reduce the workload of clinicians. Plans are under way to study the application of improved models in the diagnosis of other eye diseases and to improve models for multiclass recognition.

Funding Statement: This research was supported by The People's Republic of China Ministry of Science and Technology [2018YFF0213606-03 (Mu Y., Hu T. L., Gong H., Li S. J., and Sun Y. H.) <http://www.most.gov.cn>], the Science and Technology Department of Jilin Province [20160623016TC, 20170204017NY, 20170204038NY, 20200402006NC (Hu T. L., Gong H., and Li S.J.) <http://kjt.jl.gov.cn>], and the Science and Technology Bureau of Changchun City [18DY021 (Mu Y., Hu T. L., Gong H., and Sun Y. H.) <http://kjj.changchun.gov.cn>].

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] A. Bessaid, A. Feroui and M. Messadi, "Detection of blood vessels from retinal images using watershed transformation," *Journal of Mechanics in Medicine and Biology*, vol. 9, no. 4, pp. 633–642, 2011.
- [2] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua and S. Sivaprasad, "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network," *Journal of Computational Science*, vol. 20, no. 4, pp. 70–79, 2017.
- [3] P. Jan, M. Jörg, B. Rüdiger, H. Joachim and M. Georg, "Automated quality assessment of retinal fundus photos," *International Journal of Computer Assisted Radiology and Surgery*, vol. 5, no. 6, pp. 557–564, 2010.
- [4] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *British Journal of Ophthalmology*, vol. 90, no. 3, pp. 262–267, 2006.
- [5] A. Kotecha, S. Longstaff, A. Azuara-Blanco, J. F. Kirwan, J. E. Morgan *et al.*, "Developing standards for the development of glaucoma virtual clinics using a modified Delphi approach," *British Journal of Ophthalmology*, vol. 102, no. 4, pp. 531–534, 2018.
- [6] A. Ryo, M. Hiroshi, I. Aiko and A. Makoto, "Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier," *Ophthalmology*, vol. 123, no. 9, pp. 1974–1980, 2016.
- [7] H. Muhammad, T. J. Fuchs, N. D. Cuir, C. G. D. Moraes, D. M. Blumberg *et al.*, "Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects," *Journal of Glaucoma*, vol. 26, no. 12, pp. 1086–1094, 2017.
- [8] G. Varun, P. Lily, C. Marc, M. C. Stumpe, W. Derek *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [9] Z. X. Li, Y. F. He, K. Stuart, W. Meng, R. T. Chang *et al.*, "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," *Ophthalmology*, vol. 125, no. 8, pp. 1199–1206, 2018.
- [10] H. Z. Fu, J. Cheng, Y. W. Xu, C. Q. Zhang, D. W. K. Wong *et al.*, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2493–2501, 2018.
- [11] L. Giancardo, M. D. Abramoff, E. Chaum, T. P. Karnowski, F. Meriaudeau *et al.*, "Elliptical local vessel density: A fast and robust quality metric for retinal images," in *2008 30th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Vancouver, BC, Canada: IEEE, pp. 3534–3537, 2008.
- [12] M. Juneja, S. Singh, N. Agarwal, S. Bali, S. Gupta *et al.*, "Automated detection of Glaucoma using deep learning convolution network (G-net)," *Multimedia Tools and Applications*, vol. 79, no. 21–22, pp. 15531–15553, 2020.
- [13] H. Fan, D. Behdad and M. T. H. R. Bart, "Artery/vein classification using reflection features in retina fundus images," *Machine Vision and Applications*, vol. 29, no. 1, pp. 23–34, 2018.
- [14] H. Fan, D. Behdad, T. Tao and M. T. H. R. Bart, "Retinal artery/vein classification using genetic-search feature selection," *Computer Methods and Programs in Biomedicine*, vol. 29, pp. 197–207, 2018.
- [15] D. Mahapatra, P. K. Roy, S. Sedai and R. Garnavi, "Retinal image quality classification using saliency maps and CNNs," in *International Workshop on Machine Learning in Medical Imaging*, Springer, Cham, pp. 172–179, 2016.
- [16] S. Bianco, L. Celona, P. Napoletano and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355–362, 2018.
- [17] K. Xu, L. Zhu, R. Wang, C. Liu and Y. Zhao, "SU-F-J-04: automated detection of diabetic retinopathy using deep convolutional neural networks," *Medical Physics*, vol. 43, no. 6Part8, pp. 3406, 2016.
- [18] D. Debabrata, K. Raghvendra, B. Aishik, H. D. Jude, G. Deepak *et al.*, "Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm," *Soft Computing*, pp. 1–9, 2020.
- [19] F. L. Yu, J. Sun, A. Li, J. Cheng, C. Wan *et al.*, "Image quality classification for DR screening using deep learning," in *2017 39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju, Korea (South): IEEE, pp. 664–667, 2017.

- [20] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA: IEEE, pp. 770–778, 2016.
- [21] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *2015 37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp. 715–718, 2015.
- [22] A. Li, J. Cheng, D. W. K. Wong and J. Liu, "Integrating holistic and local deep features for glaucoma classification," in *2016 38th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, USA: IEEE, pp. 1328–1331, 2016.
- [23] S. Zhou, W. I. Liang, J. G. Li and J. U. Kim, "Improved VGG model for road traffic sign recognition," *Computers, Materials & Continua*, vol. 57, no. 1, pp. 11–24, 2018.
- [24] J. Su, R. Xu, S. Yu, B. Wang and J. Wang, "Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 5, pp. 2294–2309, 2020.
- [25] J. Su, R. Xu, S. Yu, B. Wang and J. Wang, "Redundant rule detection for software-defined networking," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 6, pp. 2735–2751, 2020.
- [26] V. Sudha and T. R. Ganeshbabu, "A convolutional neural network classifier VGG-19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning," *Computers, Materials & Continua*, vol. 57, no. 1, pp. 11–24, 2021.
- [27] J. D. Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Medicine*, vol. 24, no. 9, pp. 1342–1350, 2018.
- [28] Q. Cui, S. McIntosh and H. Sun, "Identifying materials of photographic images and photorealistic computer generated graphics based on deep CNNs," *Computers, Materials & Continua*, vol. 55, no. 2, pp. 229–241, 2018.
- [29] W. Quan, K. Wang, D. Yan and X. Zhang, "Distinguishing between natural and computer-generated images using convolutional neural networks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2772–2787, 2018.