Kalpa Publications in Engineering

Volume 4, 2022, Pages 147–157

Proceedings of International Symposium on Applied Science 2021

kalpa Engineering

Classification of Stages Diabetic Retinopathy Using MobileNetV2 Model

Hoang Nhut Huynh[1,3*], Minh Thanh Do[1,3], Gia Thinh Huynh[1,3], Anh Tu Tran[2,3] and Trung Nghia Tran[1,3*]

[1] Department of Biomedical Engineering Physics, Faculty of Applied Sciences, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, ward 14, District 10, Ho Chi Minh City, Vietnam.

[2] Laboratory of General Physics, Faculty of Applied Sciences, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, ward 14, District 10, Ho Chi Minh City, Vietnam.

[3] Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam.

nhut.huynhvlkt@hcmut.edu.vn, ttnghia@hcmut.edu.vn

Abstract

Diabetic retinopathy (DR) is a complication of diabetes mellitus that causes retinal damage that can lead to vision loss if not detected and treated promptly. The common diagnosis stages of the disease take time, effort, and cost and can be misdiagnosed. In the recent period with the explosion of artificial intelligence, deep learning has become the most popular tool with high performance in many fields, especially in the analysis and classification of medical images. The Convolutional Neural Network (CNN) is more widely used as a deep learning method in medical imaging analysis with highly effective. In this paper, the five-stage image of modern DR (healthy, mild, moderate, severe, and proliferative) can be detected and classified using the deep learning technique. After cross-validation training and testing on the corresponding 5,590-image dataset, a pre-MobileNetV2 training model is proposed in classifying stages of diabetic retinopathy. The average accuracy of the model achieved was 93.89% with the precision of 94.00%, recall 92.00% and f1-score 90.00%. The corresponding thermal image is also given to help experts for evaluating the influence of the retina in each different stage.

Keywords: diabetic retinopathy, deep learning, MobileNetV2, balancing dataset

* Corresponding author

1 Introduction

Nowadays, healthcare becomes one of the top concerns, and the early detection and early treatment of diseases become crucial. Diabetes is a disease that increases the amount of glucose in the blood caused by a lack of insulin [1]. Diabetes affects organs such as the retina, heart, nerves. Diabetic Retinopathy (DR) is a complication of diabetes that causes the blood vessels of the retina to swell and to leak fluids and blood [2]. The patient can lose sight if the disease changes badly. Retinal screening is necessary for diabetic patients for timely treatment at an early stage. The stages of the disease consist of four stages: mild nonproliferative, moderate nonproliferative, severe nonproliferative, proliferative.

In the first stage, there will be balloon-like swelling in small areas of the blood vessels in the retina. In the second stage, known as moderate nonproliferative retinopathy, some of the blood vessels in the retina will become blocked. In the third stage, severe nonproliferative retinopathy brings with it more blocked blood vessels, which leads to areas of the retina no longer receiving adequate blood flow. Without proper blood flow, the retina can't grow new blood vessels to replace the damaged ones [3]. The fourth stage or the final stage is known as proliferative retinopathy. This is the advanced stage of the disease. Additional new blood vessels will begin to grow in the retina, but they will be fragile and abnormal. Because of this, they can leak blood which will lead to vision loss and possibly blindness. Diagnosis using the naked eye is prone to misdiagnosis and requires more experience and effort. Automated diagnostic methods may save more cost and time and effective than normal diagnosis methods.

Recent research evaluating DR automated method uses deep learning to detect and classify DR. Liu et al. created a weighted paths CNN called WP-CNN to classify referable DR images in a private dataset with the accuracy (ACC) of 94.23% [4]. Das et al. proposed two independent CNNs to classify the images into normal or DR images with the ACC of 98.7% on the DIARETDB1 dataset [5]. Kassani, S.H et al. created a weighted paths CNN to classify referable DR images in an APTOS 2019 with the ACC of 83.09% [6]. Mobeen-ur-Rehman et al. proposed a simple CNN to detect the DR stages of the Messidor dataset with an excellent ACC of 98.15% [7],[8]. Zhang et al. proposed a method to detect the DR stages of their private dataset with the ACC of 96.5% [9]. They fine-tuned InceptionV3, ResNet50V2, Xception, InceptionResNetV2, and DenseNets then combined the strongest CNNs [9 - 14]. Zhang, W.; Zhong et al proposed four independent CNNs to classify the images' private dataset with the ACC of 96.5% [27].

In this study, the MobileNetV2 model is proposed in classifying the stages of diabetic retinopathy with four disease stages and healthy cases using pre-training models.

2 Methodology

2.1 Dataset

The dataset that is used in this study is an open-source dataset on the Kaggle Database [15]. This is the data of an APTOS 2019 Blindness Detection competition on Kaggle. A large dataset is provided with a set of retinal images taken using fundus imaging under different conditions. Doctors will evaluate each image of the severity of diabetic retinopathy on a scale of 0 to 4: 0 – No_DR, 1 - Mild, 2 - Moderate, 3 - Severe, 4 - Proliferative DR. Like any dataset collected in a real-world environment, it is going to be had a disturbance. Images may contain false, out-of-focus, low-light, or oversized information. Images collected from many different clinics using a variety of imaging equipment over long periods will create many other variations of the input images.

The dataset used included 5,590 images corresponding to five different stages of diabetic retinopathy. This dataset consisted of 2,409 No_DR, 744 mild, 1,511 moderate, 401 severe, and 525 cases of proliferating DR. 60% of datasets are used for training, 20% for validating, and 20% for testing. All shapes are resized to a size of 224× 224 pixels to fit the model. Table 1 presents the number of images for training, validating, and testing.

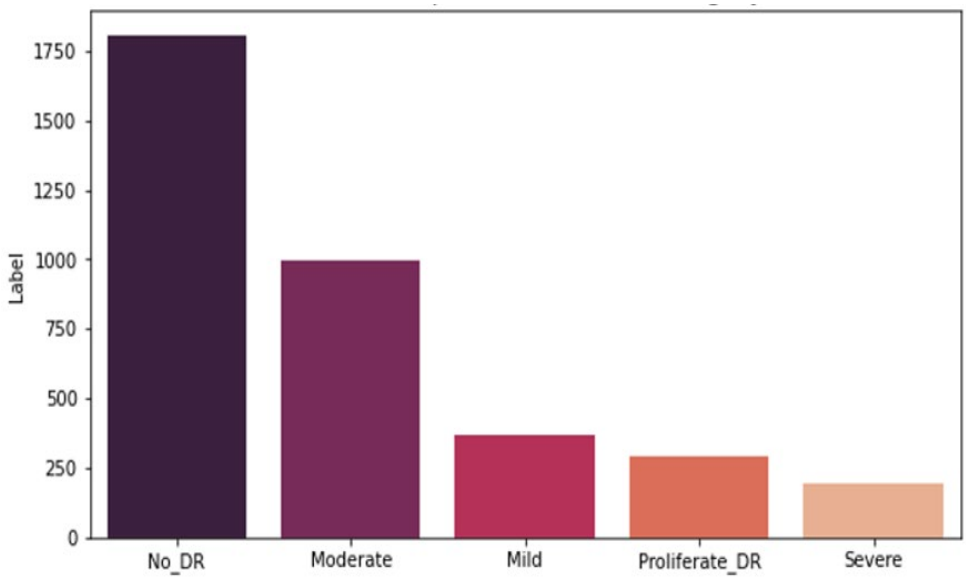

Figure 1: Training data

Figure 1 represents training data corresponding to the five stages of diabetic retinopathy. The illustration images of each stage are shown Figure 2.

Stage	Training	Validation	Testing	Total
No_DR	1805	302	302	2409
Mild	370	187	187	744
Moderate	999	256	256	1511
Severe	193	104	104	401
Proliferate DR	295	115	115	525

Table 1: Data distribution for training, validation and testing

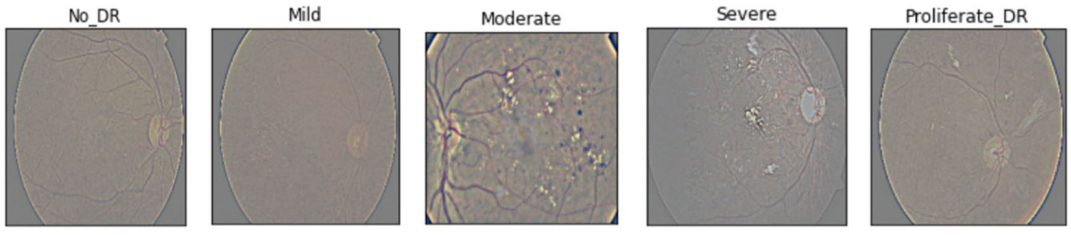

Figure 2: The illustration images of five stages in the dataset

2.2 Preprocessing Data

The training data is not balanced as shown in Figure 1. There are only 193 severe cases and 370 mild cases that are much less than the rest. If this training data will be used then the model can very well detect the cases of no_DR, moderate, and proliferate_DR than the cases of severe and mild. Therefore, the accuracy may become high but there is no balance between the layers. To overcome this problem, data enhancement techniques such as shifting, rotating, zooming are implemented so as not to lose the balance of accuracy between stages of the disease, all stages of equal importance. At the same time, the data enhancement will make the model become more general and avoid overfitting.

2.3 Convolutional Neural Networks

Deep learning (DL) is a branch of machine learning techniques that involves hierarchical layers of non-linear processing stages for unsupervised feature learning as well as for classifying patterns [16]. Nowadays, DL is one of the most effective computer-aided medical diagnosis methods [17]. The convolutional neural network (CNN) is useful in computer vision tasks. It has made impressive progress in many areas especially medical diagnostics.

Type/Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32\ dw$	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64\ dw$	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128\ dw$	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128\ dw$	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256\ dw$	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256\ dw$	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5× Conv dw/ s1	$3 \times 3 \times 512\ dw$	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512\ dw$	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024\ dw$	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$5 \times 5 \times 1024$
Softmax / s1	Classifier	$5 \times 5 \times 1024$

Table 2: The typical architecture of the MobileNetV2 network

These advances are based on the CNN network capability for extracting the features from input data sources. In this study, the network's main focus is to detect different stages of diabetic retinopathy. There are many deep learning-based methods such as restricted Boltzmann Machines, convolutional neural networks (CNNs), autoencoder, and sparse coding [19]. Some of the most well-known and most used pre-training models are VGG-16, ResNet, Xception, and MobileNet [20 - 23]. In particular, MobileNetV2 is a neural network consisting of layers that accumulate in-depth and

accumulate by point. MobileNetV2 gets the best results on the ImageNet dataset followed by the VGG-16 and ResNet50V2 [21], [24]. The input image after making our preprocessing is 224 × 224 pixels. MobileNetV2 creates a featured map on the final featured object. The number of parameters of the MobileNetV2 network is 3.3 million 7 × 7 × 1024. MobileNetV2 had computational complexity and fewer parameters by using Depth-wise Separable Convolution. Table 2 shows the typical architecture of the MobileNetV2 network.

2.4 Training Phase

For more reliable reporting, the cross-validation method is performed three times. In each stage, the training data is 80% and the validation data is 20%. Three models MobileNetV2, ResNet50V2, and VGG-16 are used for training with the same dataset. The training parameters are given in Table 3. Based on Table 3, the networks are trained using the Categorical Cross-Entropy loss function and Adam optimizer. The learning rate is set at 1e-4. The network training is carried out in 100 epochs in each stage. Since there are 03 training stages, each model is trained in 300 epochs. For the ResNet50V2 and VGG-16 models, batch size 20 is chosen because the networks have many parameter connections. The data enhancement techniques are used to enhance data and avoid overfitting. The Keras library-based neural network was deployed on a Tesla P100 GPU, 16GB of RAM provided by Kaggle [25], [26].

Training parameters	MobileNetV2	ResNet50V2	VGG-16
Input shape	$224 \times 224 \times 3$	$224 \times 224 \times 3$	$224 \times 224 \times 3$
Batch size	32	20	20
Learning Rate	1e-4	1e-4	1e-4
Optimizer	Adam	Adam	Adam
Loss function	Cross Entropy	Cross Entropy	Cross Entropy
Epoch	100	100	100
Rescale	1/255	1/255	1/255
Honrizontal flipping		Yes	Yes

Table 3: The training parameters and functions

3 Result

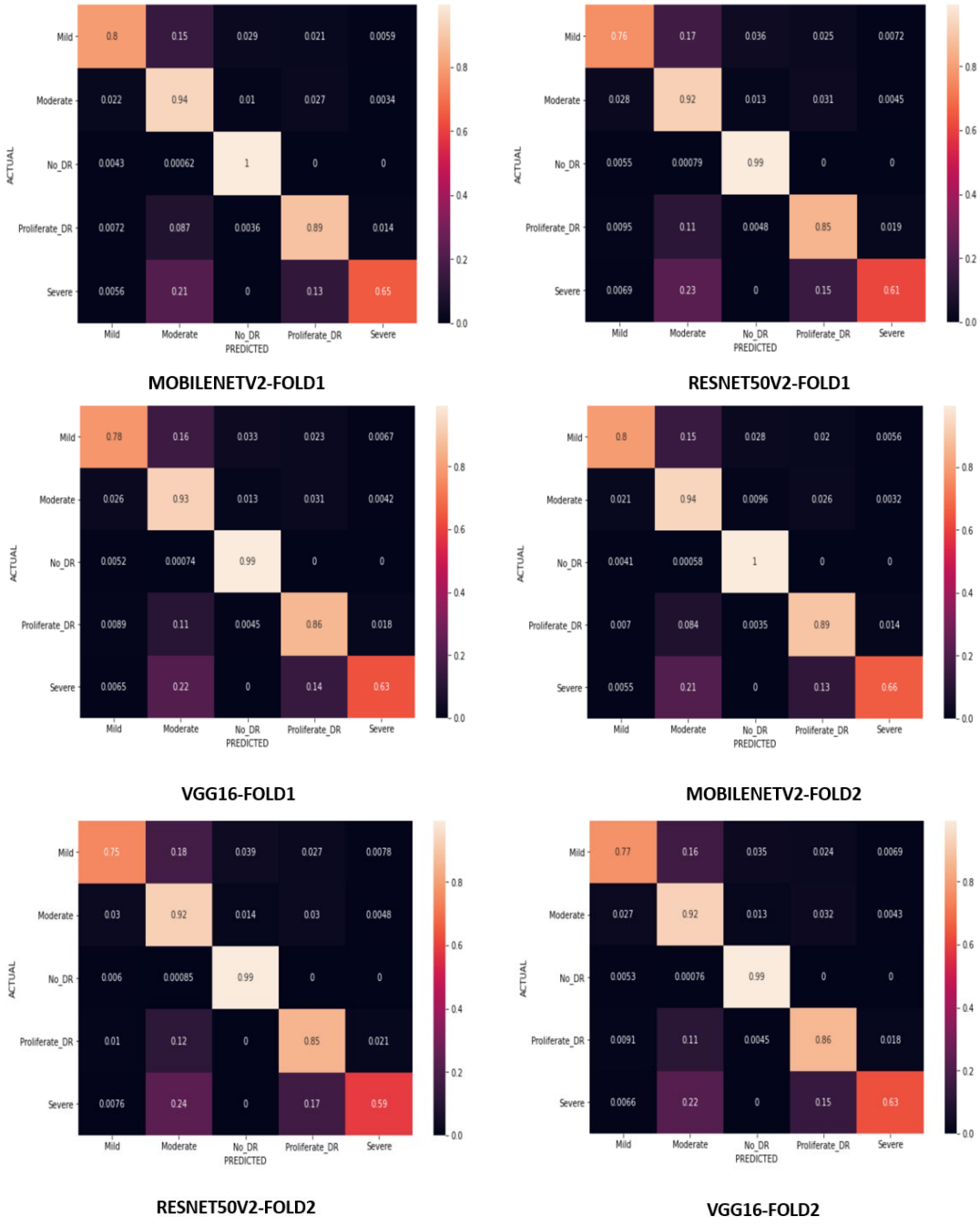

Figure 3: The confusion matrix of the networks for folds 1 and 2.

Fold	Network	Precision	Recall	f1-score	Support
1	MobileNetV2	0.94	0.91	0.91	2674
	ResNet50V2	0.88	0.88	0.87	2674
	VGG-16	0.89	0.88	0.89	2674
2	MobileNetV2	0.93	0.91	0.89	2674
	ResNet50V2	0.89	0.89	0.88	2674
	VGG-16	0.90	0.87	0.88	2674
3	MobileNetV2	0.94	0.93	0.92	2674
	ResNet50V2	0.87	0.88	0.85	2674
	VGG-16	0.88	0.84	0.85	2674

Table 4: Metrics of the model's evaluation after cross-validation

The following five-class classification model evaluation criteria:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Figure 3 shows the confusion matrix of the networks for folds 1 and 2. The validation was performed to hit the models on the dataset of 288 mild cases, 698 moderate cases, 1317 cases of no_DR, 220 cases of proliferate, and 151 severe cases. It is noteworthy that conversion learning was used during training. For training models, ImageNet weighting was used before starting training and then training based on the conditions of the dataset [24]. The accuracy is measured on the test dataset to track the performance of models for finding the best model. The results of the models are presented in Table 4.

4 Discussion

From the confusion matrix shown in Figure 3, the overall accuracy of the models is relatively high. The highest performance is the MobileNetV2 model, followed by the VGG-16 model and finally the ResNet50V2. A common feature in the results of these three models is the accuracy between layers. The No_DR are best classified with accuracy over 99%, moderate cases of accuracy over 92%, proliferate cases with over 85%. However, in mild and severe cases, the accuracy is only over 75% and 60% due to a data imbalance although the data was enhanced with the enhancement methods. A positive point here is that false positives will largely be predicted at the stage adjacent to it, with no cases of miss-predicting the stage apart. To perform performance validation of models, Table 4 shows metrics worthwhile when performing a 3-fold cross-validation method. The MobileNetV2 model has an average precision value of 94%, a recall of 92%, and an f1-score of 90%. The ResNet50V2 model has an average precision value of 88%, a recall of 88%, and an f1-score of 87%. The VGG-16 model has an average precision value of 89%, a recall of 87%, and an f1-score of 87%.

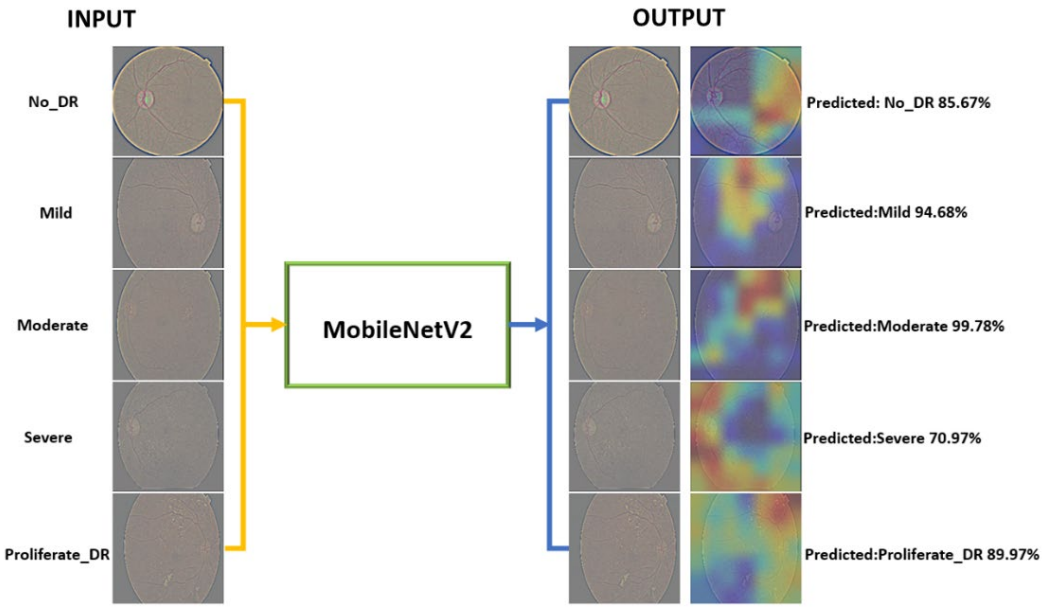

Figure 4: Image classification process: input: images of stages of DR disease; output: classification results and corresponding thermal images

From the results of the matrix confusion and cross-validation statistics, the MobileNetV2 model achieves the best classification performance in this case.

In this study, the performance of MobileNetV2 tissue in classifying the stages of diabetic retinopathy was evaluated. A total of 5,990 images was used for training the model. The overall accuracy obtained was 93.89% for the classification of five disease stages. The performance of the MobileNetV2 model is superior to the rest of the models as shown in Table 4. The highlights of this method will help classify images without the use of characteristic extraction techniques and are an effective approach that can assist diagnostic specialists. In addition, the thermal images as shown in Figure 4 can help doctors to effectively find out the zone on images of stages of diabetic retinopathy. The proposed model can diagnose the disease stages with high accuracy and high precision quickly in seconds.

One of the limitations of this method is the size of the dataset needed for training. So dataset should be needed to continuously collect and needed to be refined to remove low-quality images.

Building a new CNN architecture takes a lot of effort and time while using pre-training models that are easy to use and speed up the development process. Classifying the stages of diabetic retinopathy remains a major challenge for researchers who need more research to clarify the problem. The current work opens the way to building a complete automated monitoring system for DR which is a long-term underlying disease. The monitoring disease stage will help patients not to go blind and limit vision impairment. In our future works, YOLOv4 and YOLOv5 may be used to detect all DR lesions to obtain their benefits, such as accuracy and speed.

5 Conclusion

The incidence of diabetes is increasing worldwide, and complications of diabetic retinopathy are also increasing. Diabetic retinopathy stages are based on the type of damage that appears on the retina. This disorder threatens the vision of diabetics if diabetic retinopathy is detected in the late stages. Therefore, the detection and treatment of diabetic retinopathy in the early stages is crucial to reduce the risk of blindness. The process of diagnosing diabetic retinopathy manually with an increasing incidence of diabetic retinopathy has become insufficiently effective. Therefore, the automation of diabetic retinopathy diagnosis using a computer-aided screening system saves effort, time, and spending. Most researchers have used CNN to classify and detect diabetic retinopathy images due to its effectiveness.

In this study, the MobileNetV2 model is proposed in classifying stages of diabetic retinopathy with 93.89% accuracy with high reliability. The cross-validation was repeated three times to ensure that the model is not affected by different datasets. A thermal map is also given to help experts for evaluating the effects of different stages of the disease. The result of the proposed model shows that our model and techniques can be used to detect and classify diabetic retinopathy using deep learning. In the future, multiple datasets should be combined to achieve the balance of datasets for improving accuracy and precision.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgment

This research is funded by Ho Chi Minh City University of Technology (HCMUT) – VNUHCM under grant number SVCQ-2021-KHUD-04. We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology (HCMUT) - VNU-HCM for this study.

References

[1]. Taylor R, Batey D. Handbook of retinal screening in diabetes:diagnosis and management. second ed. John Wiley & Sons, Ltd Wiley-Blackwell; 2012.
[2]. American academy of ophthalmology-what is diabetic retinopathy? [Online]. Available, https://www.aao.org/eye-health/diseases/what-is-diabetic-retinopathy.
[3]. Grisworld Home Care Delivered with heart [Online]. Available, https://www.griswoldhomecare.com/blog/2015/january/the-4-stages-of-diabetic-retinopathy-what-you-ca/
[4]. Liu, Y.P.; Li, Z.; Xu, C.; Li, J.; Liang, R. Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. Artif. Intell. Med. 2019, 99, 101694.
[5]. 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.

[6]. Kassani, S.H.; Kassani, P.H.; Khazaeinezhad, R.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Diabetic retinopathy classification using a modified xception architecture. In Proceedings of the 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman, United Arab Emirates, 10–12 December 2019; pp. 1–6.
[7]. Mobeen-Ur-Rehman.; Khan, S.H.; Abbas, Z.; Danish Rizvi, S.M. Classification of Diabetic Retinopathy Images Based on Customised CNN Architecture. In Proceedings of the 2019 Amity International Conference on Artificial Intelligence, AICAI 2019, Dubai, United Arab Emirates, 4–6 February 2019; pp. 244–248.
[8]. Decenciere, E.; Zhang, X.; Cazuguel, G.; Lay, B.; Cochener, B.; Trone, C.; Gain, P.; Ordonez, R.; Massin, P.; Erginay, A.; et al. Feedback on a publicly distributed image database: The messidor database. Image Anal. Stereol. 2014, 33, 231–234.
[9]. Zhang, W.; Zhong, J.; Yang, S.; Gao, Z.; Hu, J.; Chen, Y.; Yi, Z. Automated identification and grading system of diabetic retinopathy using deep neural networks. Knowl. Based Syst. 2019, 175, 12–25.
[10]. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826.
[11]. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 37, 1904–1916.
[12]. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
[13]. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
[14]. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
[15]. Kaggle Database. Avalilable: https://www.kaggle.com/sovitrath/diabetic-retinopathy-224x224-gaussian-filtered
[16]. Deng Li. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans. Signal Inf Process 2014;3(2):1–29.
[17]. V Vasilakos A, Tang Y, Yao Y. Neural networks for computer-aided diagnosis in medicine : a review. Neurocomputing 2016;216:700–8
[18]. Wang X, Qian H, Ciaccio EJ, Lewis SK, Bhagat G, Green PH, Xu S, Huang L, Gao R,
[19]. Liu Y. Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction. Comput Methods Progr Biomed 2020;187: 105236.
[20]. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: a review. Neurocomputing 2016;187:27–48
[21]. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014.
[22]. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015.
[23]. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Boston, MA, 2015,: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015. p. 1–9.

[24]. Mark Sandler Andrew Howard et al., MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv:1801.04381v4 [cs.CV] 21 Mar 2019
[25]. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Ima- genet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee; 2009. p. 248–55.
[26]. Keras library. Available https://keras.io/
[27]. Kaggle. Available https://www.kaggle.com/