Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol.44 No. 6 (2023)

Diabetic Retinopathy Detection Using Neural Networks and Ensemble Learning

[1]B. Sripathy, *[2]R.Ambrose Prabhu, [1]Neha Mahesh, [1]Lalith Adhithya B

[1]*Department of Mathematics, School of Advanced Science,Vellore Institute of Technology, Vellore, 632104, Tamilnadu, India*

[2]*Department of Mathematics, Rajalakshmi Institute of Technology, Chennai, 600124, Tamilnadu, India*

*Corresponding Author

Abstract:-A very common complication of diabetes is Diabetic Retinopathy (DR) which affects the blood vessels in the retina and can result in vision loss if left untreated. For efficient treatment and the prevention of vision loss, timely identification of DR is essential. Manual retinal image screening takes time and is exorbitant, therefore, there is rising interest in automating DR screening through the use of machine learning techniques. In recent years, computerized methods for detecting DR from retinal scans have been developed using machine learning algorithms. A well-known Convolutional Neural Network(CNN) architecture is used for detecting diabetic retinopathy are discussed in this paper. The paper discusses feature extraction, classification, and image pre-processing techniques for detecting diabetic retinopathy. Finally, a comprehensive evaluation of existing approaches is presented. The challenges and opportunities for future research in this area are highlighted as well.

Keywords: *Diabetic Retinopathy, Convolutional Neural Network, image processing, Ensemble Learning*..

1. Introduction

Recent decades have witnessed a rise in the prevalence of Diabetic Retinopathy (DR), primarily as a consequence of the rise in the world's diabetes population. Diabetes is a chronic condition marked by elevated blood glucose levels, which can damage blood vessels throughout the body, including those in the retina. Another important factor influencing the rise of DR is the ageing population. DR is more prevalent as people get older since they are at an increased risk of getting diabetes and other chronic illnesses. Fundus images of the retina have become a crucial diagnostic tool for detecting DR owing to developments in medical imaging. These images can provide a non-invasive and economical way to check for DR in diabetic patients involves training a convolutional neural network model on a large dataset of labelled fundus images. Physicians in primary care and ophthalmologists can use the suggested approach as a tool to check for DR in diabetic patients, enabling early detection and decreasing the risk of blindness.Gushan[1] et al. had developed a deep learning method for early detection of DR using retainal fundus images. In their work they created convolutional neural networks architecture and classified the fundas images according to the severity of the DR, Due to COVID 19, many elderly people fear to visit hospital. Keeping this in mind, Rajalakshmi[2,] et al had developed an automated DR detection in smartphone-based fundus images using artificial intelligence. In their work they using smartphone based fundus photography at a low-cost which is accessible to everyone thus reduces the cost of travelling to visit the clinic. It has won a lot of laurels among many elderly people. Raman[3] et. al had used analyse fundus image to analyse DR using Deep learning algorithms.

Inspired by the above works Quellec[4-7] develop deep image mining for the initial screening of DR. They used Generative Adverse Networks (GAN) to detect DR . Their architecture combines a CNN as well GAN which increased the accuracy level of 0.983 for detecting DR at an early stages. Later they used machine learning methods for classification of images using SVM and random forest . They got good results in detecting DR at an early stage.

Peng[8-10] et al utilizes retinal fundus images by using 128,175 images as a data set to re classify images according to the severity. Inspired by the above works, in this paper a well CNN architecture is used for earlydetection of DR. Asiri [11] used computer-aided diagnosis system for diabetic retinopathy. One of the draw backs in this experiment is that it consumes lot of space to store the fundus image. Hence image compression plays a major role before detecting the DR. From the articles of [12-15] different authors had implemented different techniques for identification DR. This paper differs from the rest of the earlier work is that we have developed a new architecture for the detection of DR using Ensemble learning method. This method differs from the other with well known ML algorithms in detection of DR at an early stage.

2. Objectives

This article is an unsupervised learning model which takes Fundus images of the retina and predicts the existence of DR and classifies it based on the severity. We employ a CNN architecture called ResNet50, DenseNet121 and InceptionV3 to achieve our objective. The sample is trained on Nvidia RTX 3050 4GB using Google Collab. The dataset is taken from Kaggle and is called “Diabetic Retinopathy Detection”. The CNN learns to recognize the unique characteristics that differentiate objects from various classes apart, such as the object's form and texture, the presence of certain colors or patterns, and how the various parts of the thing are arranged in relation to one another. When put to the test, the trained model will automatically recognize and localize items of interest in new still photos or video frames using previously learnt attributes.

3. Methods

The dataset named “Diabetic Retinopathy Detection” is taken from Kaggle. The dataset is 88.29 GB and contains 35,126 labelled retinal images in JPEG format, including 28,126 images for training and 5,000 for testing. They were captured using a variety of cameras and imaging conditions, reflecting the diversity of real-world clinical settings. The images are labelled with a subject ID for both left and right eye images. (eg. 23_left.jpeg and 23_right.jpeg). Fig. 1 shows the left and right human eye captured through cameras.

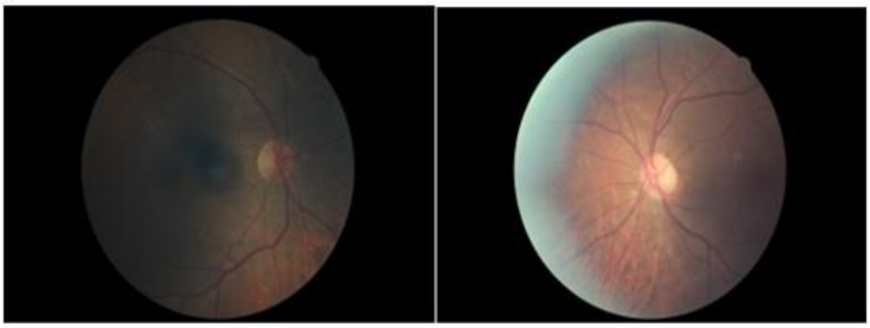

Figure 1. The left and right human eye captured through camera .

The presence of diabetic retinopathy has been rated from 0 to 4 with the scales being:

0- No DR

1- Mild DR

2- Moderate DR

3- Severe DR

4- Proliferative DR

Figure two represents the various DR images taken from the dataset.

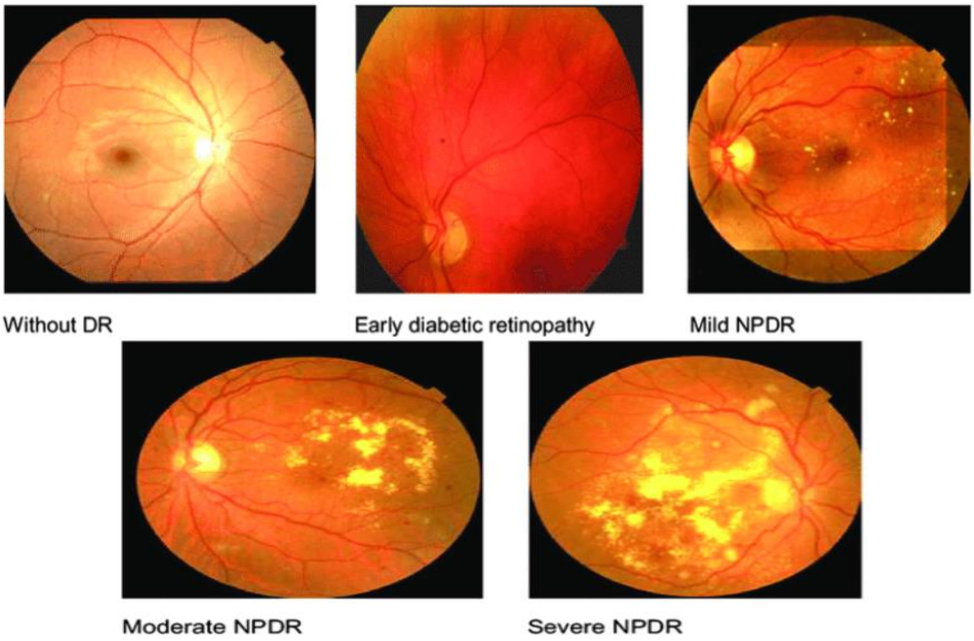

Figure 2: Different Severity of DR in a human eye .

Proliferative DR is during the final stage where the condition has gotten severe. During this stage the oxygen supply to the retina gets cut and a gel like, slimy fluid fills the back of the eye. The damages blood vessels close and cause new abnormal blood vessels to grow which can cause great threat to the retina. Severe proliferative diabetic retinopathy can lead to blindness.

Convolutional Neural Network (CNN):

CNN belongs to the deep neural network class. It processes its input data through a number of convolutional layers that apply filters and extract features from the data. These characteristics are subsequently transmitted via one or more fully connected layers, which use the extracted features to perform classification or regression tasks. One of CNNs' main advantages is their capacity to automatically learn hierarchical representations of the input data, which makes them ideal for tasks like segmentation, object recognition, and image classification. The overall architecture of the functioning of CNN was displayed in fig.3

ResNet50 Architecture:

ResNet50 is a deep neural network architecture that was introduced by Microsoft Research in 2015. It stands for "Residual Network". 50 layers constitute the ResNet50 architecture, which includes convolutional, pooling, and fully connected layers. Shortcut connections, which enable the network to skip one or more layers and send the input directly to a subsequent layer, are also included in the architecture. As a result, the problem of vanishing gradients is reduced. Large image datasets were used to pre-train ResNet50.

The issue of disappearing gradients, which is common in deep neural networks, was the primary impetus for the development of ResNet-50. This was done in the hopes of finding a solution to the issue. The gradients have a tendency to grow quite small as the network gets deeper, which makes it harder for the network to learn in an efficient manner. This problem is tackled head-on by ResNet-50's residual learning strategy, which does so by using skip connections, also referred to as shortcut connections or identity mappings.

Let's delve even further into the structure of ResNet-50's architecture:

1. Input and Preprocessing: ResNet-50 requires an input image with dimensions of 224 by 224 pixels to function properly. In typical picture preprocessing, the mean pixel values are subtracted from the total, and the image is then divided by the standard deviation.

2. Convolutional Layers: - The initial layer is a conventional convolutional layer that accomplishes a 7x7 convolution using 64 filters and a stride of 2.

A batch normalization layer and a ReLU activation come next after this step, respectively.

- After that, the output is sent through a max-pooling layer that has a stride of 2 and a window size of 3 by 3, which together bring the spatial dimensions down.

3. Residual Blocks: The ResNet-50 algorithm is divided into four stages, and each step has its own collection of residual blocks. In each of the four stages, the number of spatial dimensions is gradually reduced while the number of filters is steadily increased.

- Every one of ResNet-50's residual blocks shares the same basic structure. It begins with a convolution of size 1x1, which performs dimensionality reduction, and is then followed, by three convolutions of size 3x3, one after the other.

- After each and every convolution, batch normalization is performed, and ReLU activation is performed after the first and third convolutions.

- In order to create the residual connection, the output of the final convolutional layer in the block is combined with the initial input of the block, also known as the skip connection.

- A ReLU activation is applied on the total of the input and the residual connection in order to obtain the output of the residual block. This allows one to retrieve the output of the residual block.

4. Classification Head: - Following the application of the final residual block, a global average pooling layer is used to bring down the spatial dimensions to 1x1.

- After that, the features that were generated are sent through a fully linked layer that contains a thousand units (which correspond to the thousand ImageNet classes).

- In the final step, softmax activation is utilized so that class probabilities may be calculated.

The ResNet-50 model has a total of fifty layers, including convolutional layers, batch normalization layers, and fully connected layers. This number includes the total number of layers. The gradients are allowed to flow straight through the identity mapping thanks to the skip connections, which makes it possible for deep neural networks to be effectively trained.

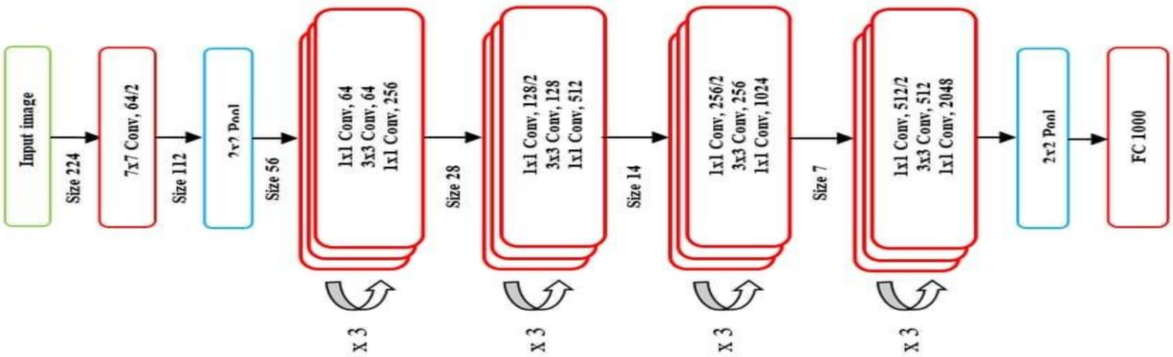

Figure 3: The architecture of CNN

The high-frequency information from the preceding layers is preserved thanks, in part, by the skip connections, which is something that is absolutely necessary for making correct predictions.

Because of its amazing performance on a variety of benchmark datasets, ResNet-50 has been a popular option for transfer learning when it comes to computer vision applications. Because of its deep design and skip connections, it is much simpler to train deeper networks and attain results in image recognition tasks that are on par with the best currently available.

DenseNet121 Architecture:

Convolutional neural network architecture known as DenseNet-121 was first presented by Huang et al. in their 2017 research titled "Densely Connected Convolutional Networks." DenseNet121 is a very efficient and densely connected architecture. The number 121 signifies the number of layers with trainable weights. DenseNet121 optimises the gradient flow during training and enables the network to learn more effectively by giving each layer access to the features of all preceding layers.

The 'vanishing gradient' issue emerges as the CNN has more layers. This means that as the channel for information from the input to the output layers increases, it may cause some information to 'vanish' or get lost, which hinders the network's capacity to train efficiently. DenseNet addresses this issue by changing the normal CNN architecture to connect each layer directly to every other layer.

Let's investigate the architecture of ResNet-50 even further by going even deeper into its structure:

1. Input and Preprocessing: The DenseNet-121 algorithm requires an input image with dimensions of 224 by 224 pixels. When preprocessing an image, it is common practise to begin by subtracting the mean pixel values from the image and then dividing the result by the standard deviation.

2. The Initial Convolutional Layer Is Comprised Of: - The initial layer of the DenseNet-121 architecture is a conventional convolutional layer that does a 7x7 convolution using 64 filters and a stride of 2. After that, the output goes via a batch normalization layer before being activated by a ReLU. - Following the completion of this preliminary convolution, the spatial dimensions are shrunk by a factor of 2 using a 3x3 max-pooling layer that has a stride of 2.

3. Dense Blocks: The DenseNet-121 structure is made up of a total of four dense blocks. Each dense block is made up of several layers, which are collectively referred to as dense layers, and is densely coupled to all of the other levels.

- The input of a layer is a concatenation of the feature maps from all of the layers that came before it in the dense block. This occurs within each dense block. Because of this high connection, information is able to flow directly from the early layers to the later layers, which encourages the reuse of features and gradient flow.

- A batch normalization layer, a ReLU activation, and a 3x3 convolution with a predetermined number of filters are typically what makes up each dense layer.

- The network is encouraged to learn representations that are both more sophisticated and richer as the number of filters contained inside each dense block gradually grows over time.

4. Transition Blocks: - There is a transition block placed in between each pair of consecutive dense blocks.

This transition block is responsible for performing dimensionality reduction and downsampling in order to regulate the expansion of feature maps and lower the amount of computational complexity.

A batch normalization layer, a 1x1 convolution layer to lower the total number of filters, and a 2x2 average pooling layer with a stride of 2 are included in the transition block. These layers are combined with a stride of 2 to bring down the total number of spatial dimensions.

5. Classification Head: - Following the application of the last dense block, a global average pooling layer is used to bring down the spatial dimensions to 1x1.

- After that, the features that were generated are sent through a fully linked layer that contains a thousand units (which correspond to the thousand ImageNet classes).

- In the final step, a SoftMax activation is utilized so that class probabilities may be calculated.

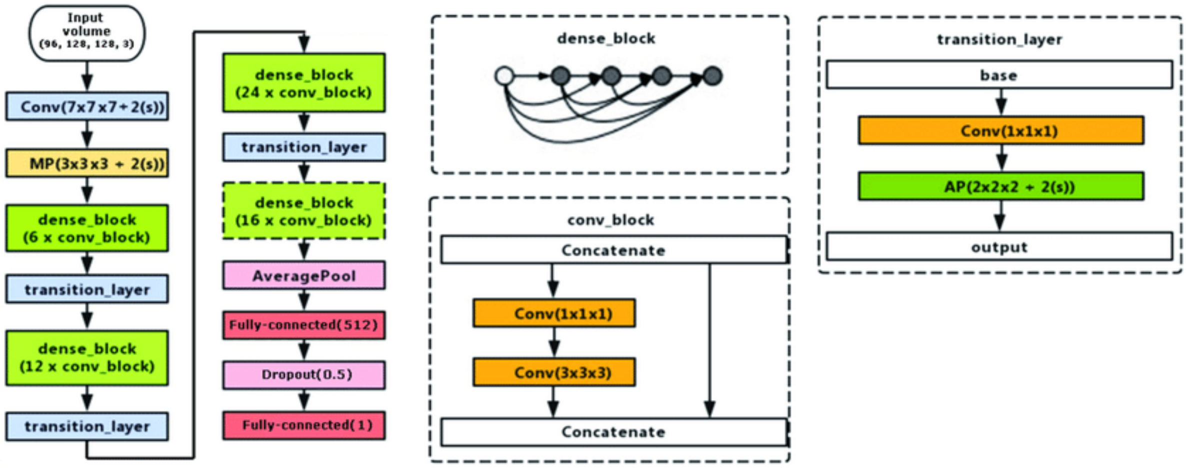

Figure 4: The architecture of DenseNet

The dense connectivity network that DenseNet-121 possesses is its primary point of differentiation. DenseNet enables feature reuse and encourages the network to learn more compact and discriminative feature representations by densely linking each layer to every other layer. This is accomplished by densely connecting each layer to every other layer. This dense connectivity helps ease the disappearing gradient problem and enhances gradient flow, which makes it easier to train deep neural networks. Additionally, it helps alleviate the difficulty of a vanishing gradient.

DenseNet-121 has shown remarkable performance on a variety of image classification benchmarks and has become a popular option for transfer learning and feature extraction in computer vision tasks as a result of these accomplishments. Its architecture encourages information flow, makes gradient propagation easier, and enables efficient learning of complicated representations, all of which contribute to state-of-the-art performance levels. DenseNet architecture was shown on fig.4

InceptionV3 Architecture:

InceptionV3 is a pre-trained CNN architecture which has 48 layers. Multiple deep layers result in complexity and results in overfitting of data. To avoid this, InceptionV3 employs numerous filters of various sizes at the same level. Therefore, adopting parallel layers in place of deep layers will result in a larger model than a deeper one. This model has a very low error rate compared to its previous models and contemporaries. This is a superior version to the previous Inception models V1 and V2.

Input Layer: This is where the input image is fed into the network. Inception V3 accepts images of size 299x299 pixels with three colour channels (red, green, blue).

Convolutional Layers: A stack of convolutional layers with tiny filter sizes (3x3 and 1x1) is the first layer in the network. These layers help with the capture of low-level elements like textures and edges.

Inception Modules: Inception V3 offers a number of layered inception modules. Each module includes various filter configurations (1x1, 3x3, 5x5) and pooling techniques to efficiently collect both local and global characteristics. The network is able to learn plenty of feature representations because of this parallel structure.

Reduction Modules: These modules are occasionally used to increase the depth of the feature maps while decreasing their spatial dimensions. This reduction aids in information compression and lowers computational complexity.

Fully linked Layers: The fully linked layers are at the very end of the network and they carry out the final classification. The high-level features that were extracted by earlier layers are mapped to the intended output classes by these levels. A softmax layer is often added towards the conclusion of Inception V3 to generate the class probabilities.

This system can take retinal images as input and provide an automated assessment of DR severity, aiding medical professionals in diagnosis and decision-making processes. The success of using Inception V3 for DR detection relies on the availability of a well-labelled and representative dataset, careful preprocessing, appropriate model configuration, and thorough evaluation.

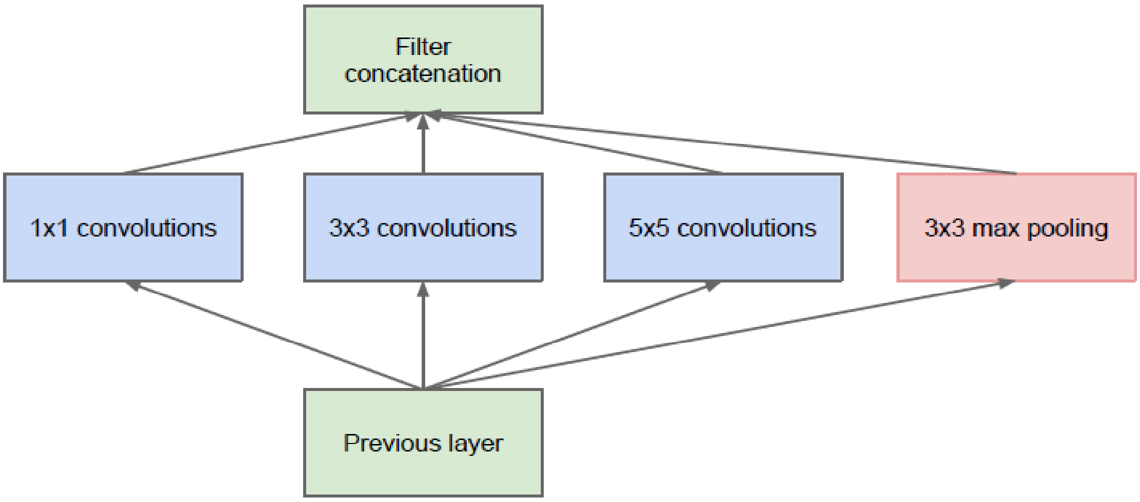

Figure 5: The Convolution used in filter processing.

Ensemble Learning:

Ensemble learning is a machine learning technique that involves combining multiple individual models, called base models, to create a more robust and accurate predictive model. By utilising the diversity and collective wisdom of various models, ensemble learning aims to increase the model's overall performance and generalization capability.

Ensemble learning operates on the principle of *"wisdom of the crowd,"* where the combined predictions of multiple models tend to be more accurate and reliable than the predictions of any individual model. The idea is that different models may make different errors or have different strengths, and by combining their predictions, the ensemble can overcome individual model limitations and produce more accurate and robust results.

In our model although the accuracy of ensemble learning is 74.08% it is observed that the error and bias of the model is lower compared to individual models.

4. Results

The training to testing ratio is taken as 80:20 to obtain optimal results. 2930 fundus images taken for training and 732 for testing have been used. All three models were run for 50 epochs. The table 1 represents the runtime and accuracy of different models . Among the different models such as ResNet50, Inception V3, DenseNet121, Ensemble models it was found that for 50 epochs ResNet 50 outperform other models in terms of Runtime and accuracy.

Table1 : Comparison of Runtime and Accuracy for different models

Model	Epochs	Run Time	Accuracy observed
ResNet50	50	25 mins	81.3%
InceptionV3	50	70 mins	75%
DenseNet121	50	41 mins	78%
Ensemble	10	21 mins	74.08%

5. Discussion

ResNet50 model outperformed InceptionV3 and DenseNet121 among the three trained and evaluated models. Considering that fact that ensemble model takes average of all three models, it has proven to be a reliable and effective model. Ensemble learning gradually improves accuracy but also it further helps reduce variance in the model. Future research can be done on the ensemble model. Ensemble model can further be used with more than 3 models to improve accuracy and reduce bias by more training and validation.

Refrences

[1] Gulshan, V, Peng. L , Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., et. al.Detection of diabetic retinopathy in retinal fundus images by developing and validating deep learning algorithm, JAMA,2016 Dec 13;316(22):2402-2410.

[2] Rajalakshmi, R., Subashini, R., Anjana, R.M. *et al.* Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye* , 2018 March 15 32(6):1138–1144

[3] Raman, R., Srinivasan, S., Virmani, S. *et al.* Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. *Eye* **2019**, April 33(6): 97–109.

[4] Nderitu, P., Nunez do Rio, J.M., Webster, M.L. *et al.* Automated image curation in diabetic retinopathy screening using deep learning. *Sci Rep* **12**, 2022, June 11196.

[5] Lam C, Yi D, Guo M, Lindsey T. Automated Detection of Diabetic Retinopathy using Deep Learning. AMIA Jt Summits Transl Sci Proc. 2018 May 18;:147-155.

[6] Chan JCN, Lim LL, Wareham NJ, Shaw JE, Orchard TJ, Zhang P et al The lancet commission on diabetes: using data to transform diabetes care and patient lives. 2019, Lancet 396(0267:2019–2082.

[7] Peng JJ, Xiong SQ, Ding LX, Peng J, Xia XB , Diabetic retinopathy: focus on NADPH oxidase and its potential as therapeutic target. Eur J Pharmacol , 2019, April 8(53):381–387.

[8] Kang EY, Hsieh Y, Li C, Huang Y, Kuo C, Kang J, Chen K, Lai C, et al.,Deep Learning–Based Detection of Early Renal Function Impairment Using Retinal Fundus Images: Model Development and ValidationJMIR Med Inform 2020;8(11):e23472

[9] Mansour, R.F. Evolutionary computing enriched computer-aided diagnosis system for diabetic retinopathy: A survey. IEEE Rev. Biomed. Eng. 2017, 10, 334–349.

[10] Asiri, N.; Hussain, M.; al Adel, F.; Alzaidi, N. Deep learning-based computer-aided diagnosis systems for diabetic retinopathy: A survey. Artif. Intell. Med. 2019, 99, 101701.