

## Diabetic Retinopathy Detection and Classification Using GoogleNet and Attention Mechanism Through Fundus Images

Amnia Salma<sup>a</sup>, Alhadi Bustamam<sup>b</sup>, Anggun Rama Yudantha<sup>c</sup>, Andi Arus Victor<sup>d</sup>,  
Wibowo Mangunwardoyo<sup>e</sup>

<sup>a,b</sup>Department of Mathematics, Universitas Indonesia, Depok, Indonesia

<sup>c,d</sup>Department of Ophthalmology, National General Hospital Dr. Cipto Mangunkusumo, Jakarta, Indonesia

<sup>e</sup>Department of Biology, Universitas Indonesia, Depok, Indonesia

Email: <sup>a</sup>amniasalma@sci.ui.ac.id, <sup>b</sup>alhadi@sci.ui.ac.id, <sup>c</sup>ramalangka@gmail.com, <sup>d</sup>arvimadao@yahoo.com,

<sup>f</sup>wibowo.mangun@ui.ac.id

Corresponding Author : Alhadi Bustamam

**Article History:** Received: XXXxx 20XX; Revised XX Xxx 20XX Accepted: XX Xxx 20XX; Published online: XXXxx 20XX

### Abstract:

Diabetic Retinopathy (DR) is one of the leading causes of blindness globally for the patient who has Diabetes. There are about 422 million people who have Diabetes in the world. Image processing in the medical field is challenging in the Big Data era. The advantages of image processing are for detection and classification of disease based on the signs of fundus images. In this research, we used the Attention Mechanism algorithm and Googlenet for detection and classification of Diabetic Retinopathy into severity levels such as normal, mild, moderate, severe, and proliferative diabetic Retinopathy. The role of attention mechanism focuses on pathological area into fundus images, and the part of Googlenet focuses on classifying fundus images into Diabetic Retinopathy levels. We used 250 datasets for training data that we obtained from Kaggle, and the accuracy of this research gets excellent performance up to 97%.

**Keywords:** Diabetic Retinopathy, Attention Mechanism, Googlenet, Fundus Images

### 1. Introduction

The World Health Organization estimates that 422 million people have Diabetes. The number of Diabetes patients is increasing significantly from year to year. The number is expected to increase to 522 million in 2034, as estimated by the International Diabetes Federation (IDF) [1]. Diabetic Retinopathy (DR), known as diabetic eye disease because it happens on retina due to Diabetes. DR is one of the leading causes of preventable blindness in the world. DR is one of the complications of Diabetes in the retina vessels [2]. Stage of DR is mild, moderate, severe, and Proliferative Diabetic Retinopathy (PDR). Every severity stage of DR has signed, such as microanarysm, cotton wool spots, exudates, hard exudates, and neovascularization [3-7]. DR is a severe disease, so suitable treatment for patients is critical to do for preventing blindness.

In the examination of the retina in conventional methods to detect DR, professional skill is needed, high cost and time consuming to perform the severity levels of DR. Doctors have to look one by one to ensure the condition of the patient and their retina to give suitable treatment for patients. From the point of healthcare view, it is more effective when DR is detected early [8].

Recent technology development in big data has enabled using artificial intelligence (AI), Machine Learning (ML), and Deep Learning (DL) for healthcare. Many approaches based on ML have been applied for detection and classification of DR in the severity levels through fundus images[9-12]. Fundus images is used for featured extraction [13]. Applied of ML used Histogram Of Oriented Gradients and Shallow learning to classify DR into Mild and Normal class achieved 85% accuracy [14].

One of the methods used in DL could improve the performance of object detection and visual object recognition [15, 16]. DL is a subset of ML that uses multiple layers containing non-linear processing units. DL application uses object detection of features of images, such as lesion features in the fundus images, for detecting DR [17].

The convolutional neural network (CNN) is one of the DL models that have convolutional, pooling, and fully connected layers. The convolutional layer extracts image features, the result of which goes to the pooling layer, where the number of feature extraction is calculated [18]. The fully connected layer is classified as a DL algorithm [19, 20]. Recent researchers using CNN for detection and classification DR. [21, 22]

CNN has been widely used for image classification [23]. One of the primary applications of CNN is in medical images. Recently, [24] used CNN to detect and classify DR through fundus images using GoogLeNet and achieved a performance of 88% for detecting normal and DR. Another recently proposed CNN-based technique for DR classification is developed by [25], where two algorithms are involved. These are GoogLeNet and transfer learning. This experiment divided DR into three levels such as No DR, mild, and Severe using GoogLeNet and transfer learning, achieving a sensitivity of 95% and a specificity of 96%. Similarly, [26] used a CNN dataset and divided it into five DR severity levels such as normal, mild, moderate, severe, and PDR. The experiment shows good performance up to 93% accuracy. They designed a CNN-based architecture that performed traffic sign classification-related tasks with a good recognition rate.

One of the algorithms that helped CNN for better accuracy to classify images is the attention mechanism (AM). The AM is an algorithm, which contains three steps for implementation, such as global, local, and fusion branches [27]. AM focuses on the pathological area after fundus images are divided into normal and DR and into four levels such as mild, moderate, severe, and PDR.

In this study, we proposed an AM and CNN as methods for the detection and classification of DR using fundus images. We used GoogLeNet as an architecture of CNN because GoogLeNet has a deep architecture. We used AM because it can focus on the pathological area, and GoogLeNet was used to classify DR into normal, mild NPDR, moderate NPDR, severe NPDR, and PDR. This research also used GoogLeNet architecture without an attention mechanism for detecting and classifying DR using fundus images. We compared the results between GoogLeNet with and without an attention mechanism.

## 2. Materials And Methods

In this section, we discuss the proposed GoogLeNet and Attention Mechanism method. Our model, GoogLeNet, and GoogLeNet with attention mechanism training 250 fundus images from the Kaggle dataset to classify normal, mild, moderate, severe, and PDR. Before training data, we did image preprocessing when the images were resized to fit GoogLeNet input image and cropped the image. By this process, we know the weight of each input. Afterward, we trained the data using GoogLeNet architecture and GoogLeNet architecture with an attention mechanism to know the classification result. Zero indicates normal fundus images, 1 indicates mild fundus images, 2 indicates moderate fundus images, 3 indicates severe fundus images, and 4 indicates PDR fundus images. The proposed method in this research used several steps. The first step is to preprocess data, including the cropping and resizing of data into a fit size. We used CNN models, such as GoogLeNet and attention mechanism algorithm, for training data. GoogLeNet has a faster performance for training data, and the attention mechanism algorithm helps the GoogLeNet to classify detailed data.

We used fundus images containing normal, mild, moderate, severe, and PDR for training data. We obtained a database from the Kaggle dataset, which consists of 250 images that we used for training the model. Each level of severity DR has a 50 fundus images dataset. We built the model to classify DR into five levels, normal, mild, moderate, severe, and PDR. Normal or No DR means that the eye is healthy and there are no DR symptoms; mild means there are microaneurysms; moderate means that the number of microaneurysms and hemorrhages is less than 20 in every quadrant as well as the presence of hard exudates; severe means that the number of microaneurysms and hemorrhages is more than 20 in more than two quadrants as well as exudates and red lesions; PDR means that neovascularization, abnormal formation of blood vessels, have developed, as evident in the fundus images.

In recent years, the development of applied AI for healthcare has been challenging for researchers. GoogLeNet are used to classify images for disease detection. In this research, we used the GoogLeNet architecture and attention mechanism algorithm to classify fundus images into DR classes based on the symptom shown. The steps of our experiment are shown in Fig.1.

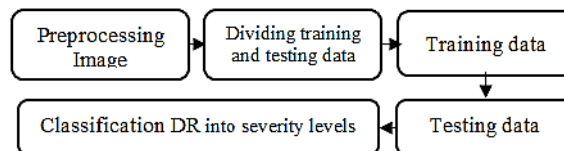


Figure 1. Flowchart DR Classification

We set the image into fit size by resizing each fundus image. Every fundus image was preprocessed, and the data was divided to 80% for training data and 20% for testing data. We used the GoogLeNet architecture and GoogLeNet architecture with an attention mechanism algorithm to train and test data. The output of this research is the classification of DR into normal, mild, moderate, severe, and PDR.

### A. Datasets and Preprocessing

We obtained a dataset from the Kaggle Dataset, which has a lot of fundus image datasets with a high resolution (<https://www.kaggle.com>). Besides that, Fundus images from Kaggle are indexed by the pathologists. The fundus images are shown in Fig.2.

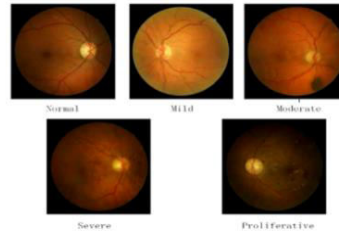


Figure 2. DR in five severe levels [28]

We provided the balance data Fundus image (normal, mild NPDR, moderate, severe, and PDR criteria shown in Fig.2). The number of each measures dataset is 50 fundus images. We have 250 fundus images from the dataset, and the data was divided into 80% for training data and 20% for validating data.

Preprocessing data is important to do. In the research, we resized the image into fit Googlenet input size 224x224x3 and cropped image. The preprocessing data are shown in Fig.3.

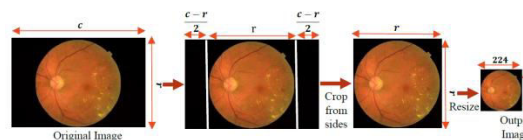


Figure 3. Preprocessing images

### B. GoogleNet

Googlenet gets popularity from different research communities and is applied for different applications. GoogLeNet is trained on ImageNet [29] with millions of high-resolution images to classify them into 1000 different classes such as a keyboard, mouse, and many species of animals with a very low error rate by tuning nearly seven million parameters doing 1.5 billion operations.

The structure inception architecture of Googlenet used some intermediate classifier branches in the middle of the architecture. These branches are used only during training the model. These branches consist of a 5x5 average pooling layer with a stride of 3, then 1x1 convolution with 128 filters, two fully connected layers of 1024 outputs and 1000 outputs, and a softmax classification layer. The generated loss of these layers added to total loss with a weight of 0.3. These layers help in combating the gradient vanishing problem and provide regularization too.

GoogleNet used 22 layers deep CNN in the 2014 ILSVRC competition, which is smaller and more speedily than VggNet, and smaller and more precise than AlexNet on the original ILSVRC images. It was designed to keep computational efficiency in the process. It can perform in low computational resources. For top-5 classification tasks, the error rate is 5.5%. The network structure is more complex than VggNet, adding 'Inception' layers to the network structure. Each 'Inception' layer contains six convolutional and one pooling operation, which decreases the thickness of the fusion feature image. The architecture of Googlenet is shown in Fig. 4.

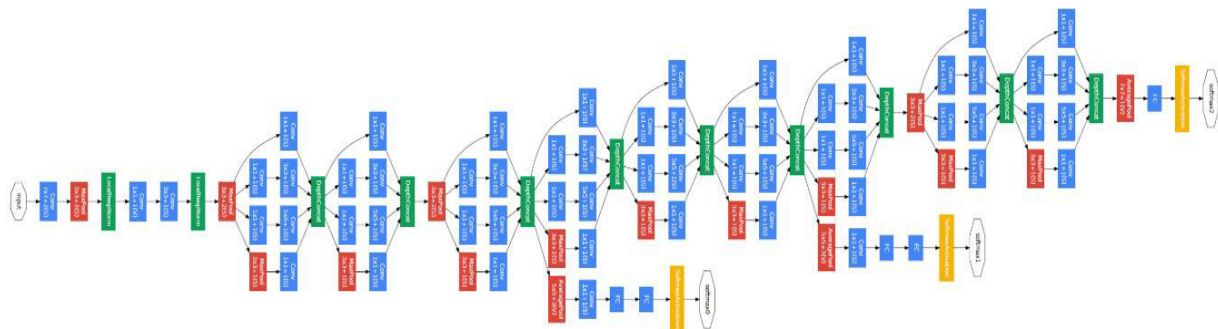


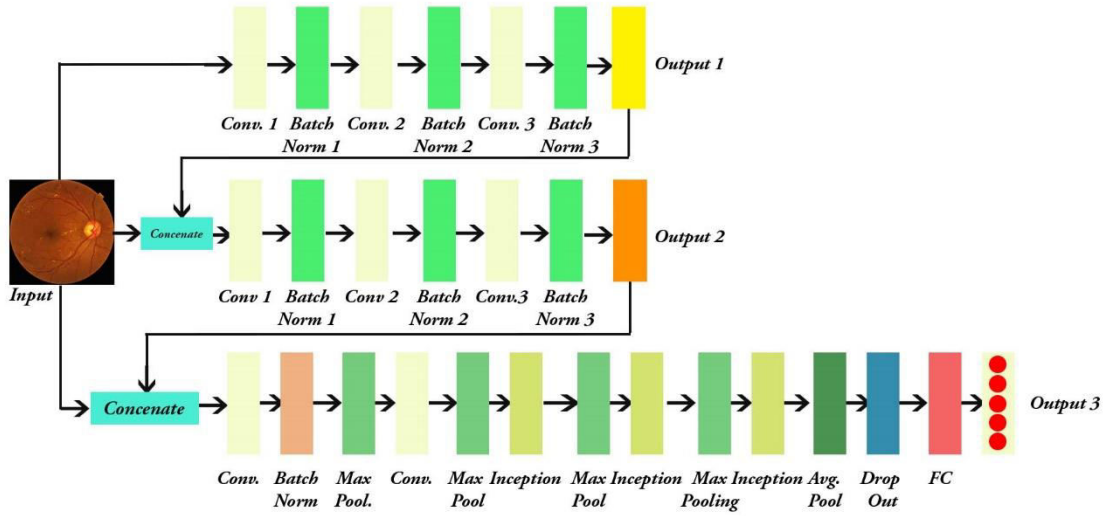
Figure 4. Goglenet architecture [29]

We used a parallel filter for operations on the input data from the previous layer, and many receptive field image sizes for convolution are, respectively, 1x1, 3x3,5x5, and pooling operation is 3x3. This architecture has two auxiliary classifier layers that are connected to the output of inception and inception layers.

GoogleNet gains the image size of 224x224 with RGB color channels the format data of images are Tiff. All the convolutions in-side the architecture use Rectified Linear Units (ReLU) as activations function. The ReLU gains the pixels of the image dataset, and if the pixels have a positive value, the value is kept, and if the value is negative, it will be converted as zero.

### C. GoogleNet and Attention Mechanism

After preprocessing data, fundus images in the first output through convolution and max pooling. The function of this process is to extract the data and determine normal and DR fundus images. After that, the part of concatenate is used to connect the result of the first output and input images into the second process, such as Googlenet architecture without a fully connected layer, and at the end process, we add deconvolution to fit the size for the next process. In the last process, we used concatenate again to connect the second process to the last process. The last process used Googlenet architecture, and we obtained the classification of multiclass.



**Figure 5. Googlenet Architecture and Attention Mechanism**

Image input for this model is 224x224x3. The first input is used to divide the fundus image into two outputs. There are normal or No DR and DR. After predicting the attention map in the second input. We design the pathological area that contains the input image and output first process. After that, the classification comes from the second output and concatenate with input image.

Attention Mechanism consists of three points, global branch, local branch, and fusion branch. First, we have to set labels for each image. Given label in every image into a vector for  $L = [l_1, l_2, \dots, l_c]$  where  $l_c \in \{0, 1\}$  and  $l_c$  represent pathological area. Global branch denoted by,

$$p_g(C|I) = \frac{1}{1 + \exp(-p(C|I))} \quad (2)$$

Where I represent global image is a probability of I for the class C. Then for parameter optimation used to a minimum of Binary cross-entropy (BCE) loss, denoted by,

$$L(W_{(g)}) = -\frac{1}{C} \sum_{c=1}^C l_c \log(P_{(g)}(c|I)) + (1-l_c) \log(1-P_{(g)}(c|I)) \quad (3)$$

Local branch denoted by,

$$L(W_{(l)}) = -\frac{1}{C} \sum_{c=1}^C l_c \log(P_{(g)}(c|I)) + (1-l_c) \log(1-P_{(g)}(c|I_c)) \quad (4)$$

And then Fusion Branch denoted by,

$$L(W_f) = -\frac{1}{C} \sum_{c=1}^C l_c \log(P_f(c|[I, I_c])) + (1-l_c) \log(1-P_f(c|[I, I_c])) \tag{5}$$

The attention mechanism on the Googlenet architecture is divided into three sections learning processes. In the first section, the learning process resulted from data used as input data for the second process and used to know the DR area and normal area. In the second process, we classified the DR area into several levels. All DR levels become output in the last process using softmax.

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

### 3. Result

In this research, we used Google Collaboratory and Python programming language along with Pytorch library to simulate the model with 250 fundus images obtained from the Kaggle dataset.

#### A. Googlenet

The result of the Googlenet architecture that we build in this research is shown in Table I.

**TABLE I. Architecture of Googlenet**

Layer Name	Input size	Filter Size/Stride/Padd	Pooling Size/Stride/Padd	1x1	3x3R	3x3	5x5R	5x5	Pool Proj	Output
Conv.	224x224x3	7x7x64/2/2	-	-	-	-	-	-	-	112x112x64
Max Pool	112x112x64	-	3x3/2/0	-	-	-	-	-	-	56x56x64
Conv.	56x56x64	-	-	-	64	192	-	-	-	56x56x192
Max Pool.	56x56x192	-	3x3/2/0	-	-	-	-	-	-	28x28x192
Inception (3a)	28x28x192	--	-	64	96	128	16	32	32	28x28x256
Inception (3b)	28x28x256	-	-	128	128	192	32	64	64	28x28x480
Max Pool	28x28x480	-	3x3/2/0	-	-	-	-	-	-	14x14x480
Inception (4a)	14x14x480	-	-	192	96	208	16	64	64	14x14x512
Inception (4b)	14x14x512	-	-	160	112	224	24	64	64	14x14x512
Inception (4c)	14x14x512	-	-	128	128	256	24	64	64	14x14x512
Inception (4d)	14x14x512	-	-	112	144	288	32	64	64	14x14x528
Inception (4e)	14x14x528	-	-	256	160	320	32	128	128	14x14x832
Max Pool	14x14x832	-	3x3/2/0	-	-	-	-	-	-	7x7x832
Inception (5a)	7x7x832	-	-	256	160	320	32	128	128	7x7x832
Inception (5b)	7x7x832	-	-	384	192	384	48	128	128	7x7x1024
Average Pool	7x7x1024	-	7x7/1/0	-	-	-	-	-	-	1024
Drop Out (0.2)	1024	-	-	-	-	-	-	-	-	1024
Softmax	1024	-	-	-	-	-	-	-	-	5

Input size in the googlenet is 224x224x3. The total layers that we used are 22 layers contains Convolution, Max Pooling, Inception, Average Pooling layer, Drop Out, and softmax.

Overall the googlenet model is the same with architecture which develops by [29] the point of difference is softmax. We used five outputs in the softmax that indicate about classify fundus images into five severity levels such normal, mild, moderate, severe, and PDR. Process of running time used Googlenet take 30 minutes with learning rate 0,001.

#### B. Googlenet and Attention Mechanism

The normal class indicates no disease, and the DR class indicates an infected or diseased eye. The model was pre-trained on 250 datasets until it reached a significant level.

Image input for Googlenet and attention mechanism model is 224x224x3. We divided the models into three sections. The first section consists of convolution layers and batch normalization. This process is used to divide the fundus image into two outputs. There are normal or No DR and DR.

In the second input, we design the model to focus on the pathological area. This section contains the input image and output first process. This process contains a convolutional layer, batch normalization, and deconvolutional layer. This process divided DR into four levels such mild, moderate, severe, and PDR.

The last process in the model is classification. Input comes from the second output and Concatenate with the input image. This process contains several layers such as a convolutional layer, max pooling, inceptions, average max pooling, drop out and softmax in the end. Output from this process is classification DR into five levels such normal, mild, moderate, severe, and PDR. The Concatenate function connects every process in the model.

Concatenate function is one of the challenging parts of the model. We have to set the first image input and the first image output suitable with the second image input for good performance in this model. The highlight of Concatenate is noticed in the weight of every input and output image. The difference in input and output images' weight will make the model getting an error.

**TABLE II. The Architecture of Googlenet and Attention Mechanism**

Layer Name	Input size	Filter Size/Stride/Padd	Pooling Size/Stride/Padd	1x1	3x3R	3x3	5x5R	5x5	Pool Proj	Output
Conv.	224x224x3	7x7x64/2/2	-	-	-	-	-	-	-	112x112x64
Max Pool	112x112x64	-	3x3/2/0	-	-	-	-	-	-	56x56x64
Conv.	56x56x64	-	-	-	64	192	-	-	-	56x56x192
Max Pool.	56x56x192	-	3x3/2/0	-	-	-	-	-	-	28x28x192
Concatenate	-	-	-	-	-	-	-	-	-	
Conv.	224x224x3	7x7x64/2/2	-	-	-	-	-	-	-	112x112x64
Max Pool	112x112x64	-	3x3/2/0	-	-	-	-	-	-	56x56x64
Conv.	56x56x64	-	-	-	64	192	-	-	-	56x56x192
Max Pool.	56x56x192	-	3x3/2/0	-	-	-	-	-	-	28x28x192
Concatenate	-	-	-	-	-	-	-	-	-	
Conv.	224x224x3	7x7x64/2/2	-	-	-	-	-	-	-	112x112x64
Max Pool	112x112x64	-	3x3/2/0	-	-	-	-	-	-	56x56x64
Conv.	56x56x64	-	-	-	64	192	-	-	-	56x56x192
Max Pool.	56x56x192	-	3x3/2/0	-	-	-	-	-	-	28x28x192
Inception (3a)	28x28x192	--	-	64	96	128	16	32	32	28x28x256
Inception (3b)	28x28x256	-	-	128	128	192	32	64	64	28x28x480
Max Pool	28x28x480	-	3x3/2/0	-	-	-	-	-	-	14x14x480
Inception (4a)	14x14x480	-	-	192	96	208	16	64	64	14x14x512
Inception (4b)	14x14x512	-	-	160	112	224	24	64	64	14x14x512
Inception (4c)	14x14x512	-	-	128	128	256	24	64	64	14x14x512
Inception (4d)	14x14x512	-	-	112	144	288	32	64	64	14x14x528
Inception (4e)	14x14x528	-	-	256	160	320	32	128	128	14x14x832
Max Pool	14x14x832	-	3x3/2/0	-	-	-	-	-	-	7x7x832
Inception (5a)	7x7x832	-	-	256	160	320	32	128	128	7x7x832
Inception (5b)	7x7x832	-	-	384	192	384	48	128	128	7x7x1024
Average Pool	7x7x1024	-	7x7/1/0	-	-	-	-	-	-	1024
Drop Out (0.2)	1024	-	-	-	-	-	-	-	-	1024
Softmax	1024	-	-	-	-	-	-	-	-	5

The performance of Googlenet and Attention Mechanism is shown in the following table.

**TABLE III. The performance of Googlenet and Attention Mechanism.**

	<b>Googlenet</b>	<b>Googlenet and Attention Mechanism</b>
Input image size	224x224x3	224x224x3
Average Accuracy	85%	97%
Average of Training time	40 minutes	30 minutes

The input size of fundus images is the same between Googlenet architecture and Googlenet with an attention mechanism. Googlenet with attention mechanism shows better accuracy than Googlenet architecture. The difference accuracy of both models is 12%, with 85% on Googlenet architecture and 97% for Googlenet with an attention mechanism. Googlenet with attention mechanism is faster than Googlenet in terms of runtime. For epoch training, both models had 20 epochs. Every model iterates five times.

#### 4. Discussion

This study compared models for detecting and classifying DR through fundus images. The results of this study showed that Googlenet and Attention Mechanism had better performance than Googlener only. The runtime of Googlenet was 40min, which was ten times longer development time than Googlenet and Attention Mechanism (30min). However, Googlenet and Attention Mechanism showed better ( $\geq 12\%$ ) accuracy than Googlenet models. The results of this study agreed with the results of previous studies [30-32] that reported that the performance of CNN and Attention Mechanism was better than CNN methods (e.g., AlexNet and ResNet) for detecting and classifying DR by fundus images. The results of this study implied that the Attention Mechanism for knowing the pathological area is effective in helping CNN classifying DR into five levels.

#### 5. Conclusion

Diabetic Retinopathy is one of the complications of Diabetes and leading blindness because the damage in retina. An automated detection and classification of the DR level has an important. Early detection allows for suitable treatment to the patient, which is crucial because early detection can effectively prevent visual blindness. DR automatic classification of fundus images can effectively help doctors in diagnosis DR, which can improve the diagnostic efficiency. In this research, applied CNN for detecting DR with Attention Mechanism are presented to detect and classify DR through fundus image. At the same time, data preprocessing can reduce for the lack of fundus image such as the image doesn't proportion. Beside that concatenate have important role of this model. The best experimental model classification achieved very good performance up to 97% in accuracy, and our results better accuracy on DR image classification using attention mechanism than without attention mechanism.

In the future, we have plans to classify DR into five classes such as every level on DR can detect correctly and CNN trained to focus not only detect normal and DR fundus images but also the classify of DR level from mild, moderate, severe, and PDR. Furthermore, we plan to obtain a dataset from a real Indonesia screening setting, such as a hospital in Indonesia that specializes in eye treatment. The ongoing development of combining CNN with other support algorithms allows deeper networks to learn the data are getting better.

#### 6. Acknowledgment

We express our gratitude to Universitas Indonesia, Faculty of Mathematics and Natural Science, who funded this research through PDUPT grant Number: NKB-183/UN2.RST/HKP.05.00/2021.

#### References

- [1] L. Guariguata, D. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, J. Shaw. (2014). Global estimates of diabetes prevalence for 2013 and projections for 2035, *Diabetes Research and Clinical Practice* 103 (2) 137–149.
- [2] Khojasteh, P., Aliahmad, B., & Kumar, D. K. (2018). Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. *BMC Ophthalmology*, 275-288.
- [3] Indumathi, G., & Sathananthavath, V. (2019). Microaneurysms Detection for Early Diagnosis of Diabetic Retinopathy Using Shape and Steerable Gaussian Features. *Telemedicine Technologies*, 57-69.
- [4] Seoud, L., Hurtut, T., Chelbi, J., Cheriet, F., Langlois, J. M. P. (2015). *Red lesion detection using diabetic retinopathy screening*. *IEEE trans-actions on Medical Imaging*.
- [5] Jaya, T., Dheeba, J., & Singh, N. A. (2015). Detection of Hard Exudates in Colour Fundus Images Using Fuzzy Support Vector Machine-Based Expert System. *Journal of Digital Imaging*, 28(6), 761–768.
- [6] Joshi, S., & Karule, P. T. (2018). A review on exudates detection methods for diabetic retinopathy. *Jurnal Biomedicine & Pharmacotherapy*, 97 (2018), 1454-1460
- [7] H. Safi, S. Safi, A. Hafezi-Moghadam, H. Ahmadi, (2018) Early detection of diabetic retinopathy, *Survey of Ophthalmology*. 601–608.

- [8] U. Ishtiaq, S. A. Kareem, E. R. M. F. Abdullah, G. Mujtaba, R. Jahangir, H. Y. Ghafoor, (2019) Diabetic retinopathy detection through artificial intelligent techniques: a review and open issues, *Multimedia Tools and Applications* (21-22) 15209–15252.
- [9] Gulshan V, Peng L and Coram M (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs *Journal of American Medical Association*.
- [10] Lu W, Yan T, Yue Y, Yiqiao X, Changzheng C (2018) Applications of Artificial Intelligence in Ophthalmology: General Overview *Journal of Ophthalmology* 1-15.
- [11] S. Stolte, R. Fang, A. (2020) survey on medical image analysis in diabetic retinopathy, *Medical Image Analysis*.
- [12] Bustamam, A., Sarwinda, D., Abdillah, B., & Kaloka, T. P. (2020). Detecting lesion characteristics of diabetic retinopathy using machine learning and computer vision. *International Journal on Advanced Science, Engineering and Information Technology*, 10(4), 1367-1373.
- [13] D. Sarwinda, A. Bustamam and A. M. Arymurthy, "Fundus image texture features analysis in diabetic retinopathy diagnosis," 2017 Eleventh International Conference on Sensing Technology (ICST), 2017, pp. 1-5
- [14] D. Sarwinda, T. Siswantining and A. Bustamam, "Classification of Diabetic Retinopathy Stages using Histogram of Oriented Gradients and Shallow Learning," *2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, 2018, pp. 83-87
- [15] Maier, A., Syben, C., Lasser, T., & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. *ScienceDirect*, 1-16.
- [16] Abdillah, Bariqi., Bustamam, A., Sarwinda, D. (2017) *Classification of Diabetic Retinopathy Through Texture Features Analysis. ICACISIS*, 333-339.
- [17] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*.
- [18] Lecun, Y. Bengio and G. Hinton. (2015) "Deep Learning", *nature*, vol. 521, no. 7553, 436-444.
- [19] Li, Zhuo ling., Minghui Dong, Shiping Wend, Xiang Hub. Pan Zhou & Zhigang Zeng. (2019). CLU-CNNs: Object detection for medical images. *ELSEVIER*, 53-59.
- [20] Biradjar, U., Gadhav, S., Chikodikar, S., Dadich, S., Chiwhane, S., (2019) Detection and Classification of Diabetic Retinopathy Using AlexNet Architecture of Convolutional Neural Networks. *Proceeding of International Conference on Computational Science and Applications*.
- [21] Li, Liu., Mai Xu, Xiaofei Wang, Lai Jiang & Hanruo Liu. (2019). Attention Based Glaucoma Detection: A Large-scale Database and CNN Model. *IEEE Xplore*, 10571-10580
- [22] Salamat, Nadeem., Malik M. & Saad Missen, Aqsa Rashid. diabetic retinopathy techniques in retinal images: A review. (2019) *ELSEVIER*, 168-188.
- [23] Kharabe S, Nalini C. (2018) Using adaptive thresholding extraction—robust ROI localization-based finger vein authentication. *J Adv Res Dyn Control Syst*.
- [24] Levi G, Hassner T (2009) Sicherheit und Medien. Sicherheit und Medien. doi: 10.1109/CVPRW.2015.7301352
- [25] Salma, A., Bustamam, A., Sarwinda, D. (2020). *Diabetic Retinopathy Detection Using GoogleNet Architecture of Convolutional Neural Network Through Fundus Images*. NST Proceeding.
- [26] Lam, C., Yi, D., Guo, M., & Lindsey, T. (2018). Automated Detection of Diabetic Retinopathy using Deep Learning. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science, 2017*, 147–155.
- [27] Iyyanar P , Parthasarathy J. (2020). Diabetic Retinopathy Classification Using Deep Learning Framework. *Journal of Critical Reviews*. Vol 7.
- [28] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). *Going Deeper with Convolution*. ILSVRC
- [29] Qin, H., & Peng, L. (2017). Convolutional Neural Network with Attention Mechanism for Historical Chinese Character Recognition. *Proceedings of the 4th International Workshop on Historical Document Imaging and Processing - HIP*
- [30] Sanghyun Woo , Jongchan Park , Joon-Young Lee, In So Kweon (2018) *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3-19
- [31] Wang, W., & Shen, J. (2017). Deep Cropping via Attention Box Prediction and Aesthetics Assessment. *2017 IEEE International Conference on Computer Vision (ICCV)*.
- [32] Zhang, J., Bargal, S. A., Lin, Z., Brandt, J., Shen, X., & Sclaroff, S. (2017). Top-Down Neural Attention by Excitation Backprop. *International Journal of Computer Vision*, 126(10), 1084–1102.