**Dual Convolutional Neural Networks of Ensemble learning with Attention Mechanism for Classification Task**

**using Diabetic Retinopathy dataset**

**Abstract**

Diabetic Retinopathy (DR) is a common complication of diabetes that can lead to vision problems and even blindness if not detected early. Convolutional neural networks (CNN) are now the leading method for automatically detecting DR using fundus images. These networks extract important features to identify and classify retina lesions. This research will measure the effectiveness of a combined model consisting of two CNN models with an attention mechanism in detecting and classifying diabetic retinopathy. It will also be compared with other models to assess its performance.

**Keywords:** Medical Images Classification; CNN; Attention Mechanism; Ensemble learning

# **Introduction**

Diabetes and its complications are becoming increasingly common all over the world [1]. A convolutional neural network (CNN), is a deep learning-based neural network built for processing organized arrays of input in any form, such as photographs. These are widely utilized in computer vision and have been the focal point for many applications such as image processing, as well as success in natural language processing for text categorization [2]. CNN, have recently gained significant success in many fields of medical image processing. By extracting strong characteristics, these networks can recognize complicated patterns. There characteristics were retrieved using a variety of filters that took use of the data’s inherent structure [1]. This study will record the performance of the integrated model composed of two CNN models with attention mechanism in the detection and classification of diabetes retinopathy. And compare it with other models.

# **Related work**

In recent years, deep learning approaches, particularly convolutional neural networks (CNNs) incorporating attention mechanisms, have received a lot of interest in the field of medical image processing. Researchers have investigated the use of these models in a variety of medical imaging applications, including diabetic retinopathy categorization.

AL-ANTARY and ARAFA [1] introduce the multi-scale attention network (MSA-Net) for classifying diabetic retinopathy. The method utilizes an encoder network to transform the retina image into a high-level feature pyramid is employed to capture the retinal structure in various localities. Resolution features are used to enhance the representation. Finally, for the description of the retinal structure in another location a multiscale feature pyramid is introduced.

NAZIH et al. [3] introduce a new deep learning approach using Vision Transformer (ViT) for identifying the severity stages of Diabetic Retinopathy (DR) from fundus photography images. The model was trained on the fine-grained annotated diabetic retinopathy (FGADR) dataset and optimized using the AdamW optimizer to capture the overall context of the images.

Athira and Nair [4] designed an algorithm. This algorithm utilizes rich image processing techniques, automatic hyperparameter adjustment, and neural network training strategies, with a greater emphasis on small features for better prediction. The results show that the network based on ResNet50 has superior performance in detecting and classifying diabetes retinopathy after using the proposed algorithm.

Lslam et al.[5] have created a new deep learning structure called DiaNet, which can determine if a person has diabetes or not by analyzing a photograph of their retina. Despite using a limited dataset, DiaNet utilizes a multi-stage convolutional neural network and can achieve an accuracy rate of more than 84%. Additionally, DiaNet is able to identify the specific regions of the retina images that are important in its decision-making process, which has been confirmed by medical experts.

Phridviraj et al. [6] propose a bi-directional extended short-term memory-based diabetic retinopathy detection model using retinal fundus images. The model utilizes these images to detect and classify various grades of DR. Through the application of a bi-directional LSTM method, the model is able to effectively diagnose and classify diabetic retinopathy.

Adriman et al. [7] introduce a system designed to detect and classify DRs. Their methodology consists of two primary stages: firstly, they employ local binary patterns (LBP) to extract texture features, and secondly, they extensively analyze state-of-the-art deep learning techniques for the detection and classification tasks. The deep learning techniques utilized in their study include ResNet, DenseNet, and DetNet. Initial findings indicate that ResNet achieves an accuracy of 0.9635, DenseNet achieves 0.8405, and DetNet achieves 0.9399.

# **Material and Method**

In this part, we will explain the dataset and the proposed model architecture. We describe the result model’s hyperparameters and the evaluation measures that were employed. The total number of images used in this experiment is 3663. As there is no test set, data split is required. Split the data in a ratio of 70:15:15 between training set, testing set, and validation set.

## **3.1 Dataset**

The dataset includes a color image folder and a train.csv file. The color images are composed for retinal scanning images, which are used to detect diabetes retinopathy. These images are 224 × 224 pixels. According to the severity / stage of diabetes retinopathy, all images have been saved to their respective folders using the train.csv file provided. And five directories with images: No\_DR (0), Mild (1), Moderate (2), Severe (3), Proliferate\_DR (4). The numbers represent the severity of their being recorded in the train.csv file. The following will introduce datasets under different directories:

1. ***No\_DR (No Diabetic Retinopathy):*** This dataset contains eye scanning images without diabetes retinopathy. The Figure shows some randomly selected samples from the dataset:

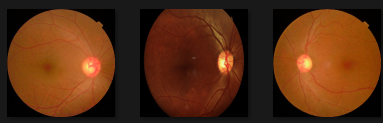


Figure 1 Some random samples in dataset No\_DR

1. ***Mild (Mild Diabetic Retinopathy):***This dataset contains eye scanning images of mild diabetes retinopathy. The Figure shows some randomly selected samples from the dataset:

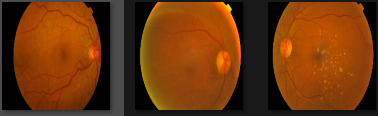


Figure 2 Some random samples in dataset Mild

1. ***Moderate (Moderate Diabetic Retinopathy):*** This dataset contains eye scans of moderate diabetes retinopathy. The Figure shows some randomly selected samples from the dataset:



Figure 3 Some random samples in dataset Moderate

1. ***Severe (Severe Diabetic Retinopathy):*** This dataset contains eye scanning images of severe diabetes retinopathy. The Figure shows some randomly selected samples from the dataset:

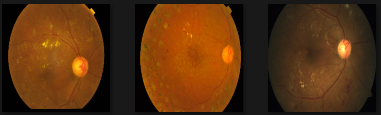


Figure 4 Some random samples in dataset Severe

1. ***Proliferate\_DR (Proliferative Diabetic Retinopathy):*** This dataset contains eye scanning images of proliferative diabetes retinopathy. The Figure shows some randomly selected samples from the dataset:



Figure 5 Some random samples in dataset Proliferate\_DR

## **3.2 Proposed Model**

This study defined a deep learning model, which is an ensemble model consisting of two different ResNet variants (Not a pre-train model). Among them, one model introduces channel attention mechanism, and the other model introduces spatial attention mechanism. The output results of these two models are added together to form the final output. The following schematic diagram shows the approximate structure of the model:

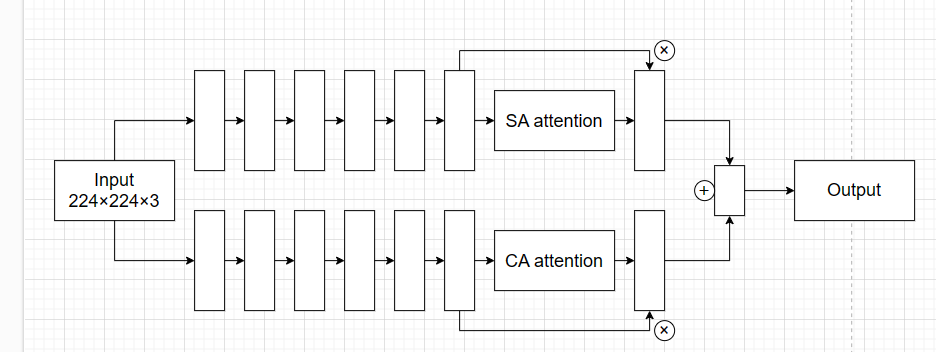


Figure 6 A schematic diagram of the approximate structure of the model

The main feature is the use of residual blocks to build the network. Residual block is a type of skip connection structure that can directly pass the input signal to the subsequent layers, thus avoiding the vanishing and exploding gradient problems.

Before introduce the CA attention and SA attention. Bakr et al. [8] introduced that attention modules are specifically engineered to eliminate unwanted noise while preserving valuable information through the process of attention scaling. This is achieved by refining the acquired features and selectively focusing on significant aspects, similar to how the human perception process operates. In this process, higher-level information guides the bottom-up learning process, enabling the capture of more intricate features while disregarding irrelevant details. The human perception and visual attention are further improved through the influence of top-down stimuli, which actively suppress non-relevant neurons through feedback loops.

CA attention: Channel attention module, used to enhance the model's attention to channel features. It includes a global average pooling layer and a global maximum pooling layer, as well as two convolutional layers and a sigmoid activation function. In the forward propagation method, it first performs global average pooling and global maximum pooling on the input, then processes and adds them through convolutional layers, and finally outputs them through the sigmoid activation function.

SA attention: The spatial attention module is used to enhance the model's attention to spatial features. It includes a convolutional layer and a sigmoid activation function. In the forward propagation method, it first performs average pooling and maximum pooling on the input, then processes and connects them through convolutional layers, and finally outputs them through the sigmoid activation function.

## **3.3 Evaluation Strategy**

These metrics used in this report include Accuracy (1), Precision (2), Recall (3), F1-Score (4) and Support. Each of the metrics is mathematically expressed as follow:

|  |  |  |
| --- | --- | --- |
| Accuracy |  | (1) |
| Precision |  | (2) |
| Recall |  | (3) |
| F1-Score |  | (4) |

Table 1 Mathematical expressions for different indicators

Accuracy: The proportion of correctly predicted samples out of the total samples, measuring the overall correctness of the model’s predictions.

Precision: The proportion of true positive predictions out of all positive predictions made by the model, measuring the accuracy of the model's positive predictions.

Recall: The proportion of true positive predictions out of all actual positive samples, measuring the model's ability to predict positive samples.

F1-Score: The weighted harmonic means of precision and recall, providing a balanced assessment of the model's performance by considering both precision and recall.

Support: The number of actual occurrences of each class in the dataset, providing insight into the distribution of different classes in the dataset.

## **3.4. Environment Execution**

The models discussed in this study were implemented using Python in a Lenovo Windows system with 64 GB of RAM. The system is equipped with an Nvidia GeForce RTX 2060 chip, known for its high-performance capabilities in handling intensive computational tasks. Additionally, the system features an Intel® Core(TM) i7-10875K CPU with a base clock speed of 2.30 GHz providing substantial processing power for the model training and evaluation processes.

# **Experimental Results**

**4.1. Performance Results using the dataset**

Table 2 shows the data of various indicators for different degrees of lesions and the overall indicator data. Figure 7 shows the confusion matrix of the test results.

Figure 8 and Figure 9 respectively show the loss and accuracy in this experimental training

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| Mild | 0.55 | 0.52 | 0.53 | 33 |
| Moderate | 0.69 | 0.76 | 0.73 | 89 |
| No\_DR | 0.96 | 0.98 | 0.97 | 162 |
| Proliferate\_DR | 0.61 | 0.54 | 0.57 | 26 |
| Severe | 0.70 | 0.41 | 0.52 | 17 |
| **Total** |  |  | **0.81** | **327** |

Table 2 Data of different degrees of lesions under various indicators and the total F1-Score

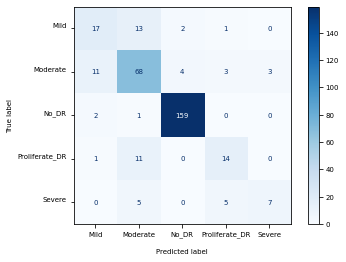
****

Figure 7 Confusion matrix

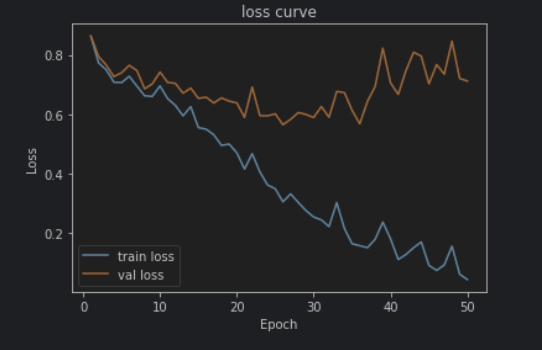


Figure 8 Model’s Loss

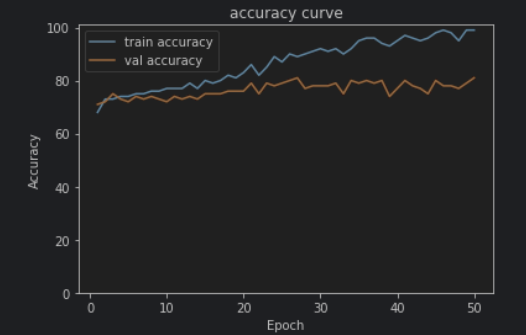


Figure 9 Model’s Accuracies

## **4.2. Discussion**

From Figure 8, it can be seen that my model is overfitting. Training should be stopped when the validation set rises, but the accuracy of such a model cannot reach a good state. This also indicates that there is still room for improvement in my model. Figure 7 shows the image of confusion matrix. By observing this image and combining it with the data in Table 2, it can be seen that the model has good detection and classification performance for images with No Diabetic Retinopathy (No\_DR) images. And performed poorly at the Severe level.

From the overall results, the value of F1-Score is 0.81. This may seem like a good value, but it also indicates that my model has many areas for improvement.

## **4.3 Fair comparison With other Deep Learning Models**

In this section, different pre trained models will be used to train the dataset. And compare the results of different pre-trained models with the results of the model proposed in this report. The results are shown in Table 6. And the results of each different pre-train model shown in Table 3, Table 4 and Table 5:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| Mild | 0.58 | 0.55 | 0.56 | 33 |
| Moderate | 0.72 | 0.74 | 0.73 | 89 |
| No\_DR | 0.96 | 0.99 | 0.98 | 162 |
| Proliferate\_DR | 0.71 | 0.58 | 0.64 | 26 |
| Severe | 0.40 | 0.35 | 0.38 | 17 |
| **Total** |  |  | **0.81** | 327 |

Table 3 resnet18 pre-train model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| Mild | 0.48 | 0.30 | 0.37 | 33 |
| Moderate | 0.58 | 0.90 | 0.70 | 89 |
| No\_DR | 0.98 | 0.99 | 0.98 | 162 |
| Proliferate\_DR | 0.67 | 0.08 | 0.14 | 26 |
| Severe | 0.00 | 0.00 | 0.00 | 17 |
| **Total** |  |  | **0.77** | **327** |

Table 4 shufflenet\_v2\_x1\_0 pre-train model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| Mild | 0.63 | 0.52 | 0.57 | 33 |
| Moderate | 0.68 | 0.75 | 0.71 | 89 |
| No\_DR | 0.96 | 0.98 | 0.97 | 162 |
| Proliferate\_DR | 0.59 | 0.65 | 0.62 | 26 |
| Severe | 0.43 | 0.18 | 0.25 | 17 |
| **Total** |  |  | **0.80** | **327** |

Table 5 mobilenet\_v3\_small pre-train model

|  |  |  |
| --- | --- | --- |
|  | F1-Score | Support |
| Resnet18 | 0.81 | 327 |
| mobilenet\_v3\_small | 0.80 | 327 |
| shufflenet\_v2\_x1\_0 | 0.77 | 327 |
| Our Model | 0.81 | 327 |

Table 6 Comparison F1-Score with other Deep Learning Models

Form the data in the above Table 3, it can be seen that there is no difference in the F1 Score between the model proposed in this report and other pre trained models. Even better than some pre trained models.

In order to ensure fairness, the computer configurations used for each model are the same, and the dataset used for this comparison is the same. However, by observing the results of each indicator of each pre trained model, it can still be seen that the model proposed in this report has very good performance

## **4.4 Indirect comparison With Existing Literature**

In this section will compare some work on diabetes retinopathy in the literature. Li et al. [9] employed a variety of pretrained CNN models, including AlexNet, GoogleNet, and VGGNet. They attained an AUC of 98.34%, an accuracy (Acc) of 92.01%. From the results, their accuracy is higher. But they are using pre-train models with pre-train parameters. And the model I am using is a newly constructed one, which has certain disadvantages. Monteiro [10] used ten state-of-the-art deep learning models, and suggested a hybrid deep learning model by merging such predictions into final score. From the results, Monteiro’s [10] model has a significant difference in accuracy for different degrees of lesions. For the classification of images with all degrees of lesions, his model has an accuracy rate of only 64.5%. From the results alone, my model has an accuracy of 81%, which is superior to his hybrid model. Sousa and Camilo [11] proposes a binary hierarchical combination of four convolutional neural networks, and the proposed technique has an accuracy of 0.853 in testing. The main process of their model is: Preprocess the images and then put them together into Model A, where Model A filters out No\_DR images. Then put it into Model B, continue to filter images of other degrees of lesions, and so on. Overall, their model is more complex than mine. But according to the data they published in the article, not all four models can achieve an accuracy of 0.853. For the results alone, their accuracy is superior to mine. But this may be largely due to Sousa and Camilo proposed [11] the hierarchical binarization of the models. Which allows even if the images do not match, they will be classified into classes closer to the correct class.

To sum up of the indirect comparison with other existing literature. My model is simpler in structure compared to other models in the literature. Despite achieving good accuracy, it may not be sufficient due to the limited structural hierarchy. To improve accuracy, more training data and a more complex model may be necessary.

# **Conclusion**.

In this report, we propose an integrated model. This is a new model framework for detection and classification, which exhibits excellent performance in detection and classification, and can effectively detect lesions and classify them. By introducing attention mechanism, the model is able to focus on key areas when processing images, improving the accuracy and performance of the model. In order to further improve the performance of the model, it is possible to consider adding more data samples in the future, especially covering images and data of different severity levels. Because this will help improve accuracy and has a wide range of applications in the medical field.

**References**

[1] M. T. Al-Antary and Y. Arafa, ‘Multi-Scale Attention Network for Diabetic Retinopathy Classification’, *IEEE Access*, vol. 9, pp. 54190–54200, 2021, doi: 10.1109/ACCESS.2021.3070685.

[2] K. Sangeetha, K. Valarmathi, T. Kalaichelvi, and S. Subburaj, ‘A broad study of machine learning and deep learning techniques for diabetic retinopathy based on feature extraction, detection and classification’, *Measurement: Sensors*, vol. 30, p. 100951, Dec. 2023, doi: 10.1016/j.measen.2023.100951.

[3] W. Nazih, A. O. Aseeri, O. Y. Atallah, and S. El-Sappagh, ‘Vision Transformer Model for Predicting the Severity of Diabetic Retinopathy in Fundus Photography-Based Retina Images’, *IEEE Access*, vol. 11, pp. 117546–117561, 2023, doi: 10.1109/ACCESS.2023.3326528.

[4] T. R. Athira and J. J. Nair, ‘Diabetic Retinopathy Grading From Color Fundus Images: An Autotuned Deep Learning Approach’, *Procedia Computer Science*, vol. 218, pp. 1055–1066, Jan. 2023, doi: 10.1016/j.procs.2023.01.085.

[5] ‘DiaNet: A Deep Learning Based Architecture to Diagnose Diabetes Using Retinal Images Only | IEEE Journals & Magazine | IEEE Xplore’. Accessed: Dec. 14, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9328261

[6] M. S. B. Phridviraj, R. Bhukya, S. Madugula, A. Manjula, S. Vodithala, and M. S. Waseem, ‘A bi-directional Long Short-Term Memory-based Diabetic Retinopathy detection model using retinal fundus images’, *Healthcare Analytics*, vol. 3, p. 100174, Nov. 2023, doi: 10.1016/j.health.2023.100174.

[7] R. Adriman, K. Muchtar, and N. Maulina, ‘Performance Evaluation of Binary Classification of Diabetic Retinopathy through Deep Learning Techniques using Texture Feature’, *Procedia Computer Science*, vol. 179, pp. 88–94, Jan. 2021, doi: 10.1016/j.procs.2020.12.012.

[8] E. M. Bakr, A. El-Sallab, and M. Rashwan, ‘EMCA: Efficient Multiscale Channel Attention Module’, *IEEE Access*, vol. 10, pp. 103447–103461, 2022, doi: 10.1109/ACCESS.2022.3205602.

[9] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, ‘Automatic Detection of Diabetic Eye Disease Through Deep Learning Using Fundus Images: A Survey’, *IEEE Access*, vol. 8, pp. 151133–151149, 2020, doi: 10.1109/ACCESS.2020.3015258.

[10] F. C. Monteiro, ‘Diabetic Retinopathy Grading using Blended Deep Learning’, *Procedia Computer Science*, vol. 219, pp. 1097–1104, Jan. 2023, doi: 10.1016/j.procs.2023.01.389.

[11] T. F. de Sousa and C. G. Camilo, ‘HDeep: Hierarchical Deep Learning Combination for Detection of Diabetic Retinopathy’, *Procedia Computer Science*, vol. 222, pp. 425–434, Jan. 2023, doi: 10.1016/j.procs.2023.08.181.