Kathford Journal of Engineering and Management,
ISSN: 2661-6106.
© 2024 *Kathford Journal of Engineering and Management*
https://kathford.edu.np/Journal/

VOL. 4 ISSUE 1 DECEMBER 2024, PP. 01-20

Deep Learning-Based Analysis for Diabetic Retinopathy Identification

Ayush Pandey[1,*], Ayushma Pandey[1], Kenish Maharjan[1], Kiran Shrestha[1], Pranita Upadhyaya[1]

[1]*Department of Computer & Electronics, Communication & Information Engineering, Kathford International College of Engineering and Management (Affiliated to Tribhuvan University), Balkumari, Lalitpur, Nepal*

*Corresponding author: *ayushpandey.076@kathford.edu.np*

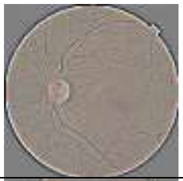
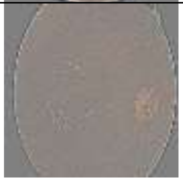
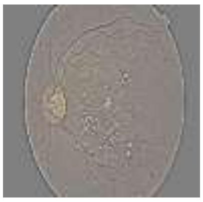
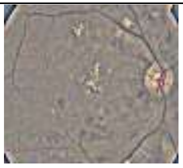
ABSTRACT— The eye condition known as diabetic retinopathy (DR) is characterized by damage to the blood vessels in the retina. Blindness may result if it is not identified in a timely manner. Early detection and treatment of DR can greatly lower the risk of visual loss. Experts with a great deal of training often use colored fundus photos to diagnose this terrible disease. Compared to computer-aided methods, manual diagnosis of DR retina fundus images by ophthalmologists takes longer because of the rising number of diabetic patients worldwide. Consequently, automatic DR detection is becoming essential. With an emphasis on medical research, deep neural network applications in healthcare have advanced significantly. The goal of this effort is to identify the five stages of DR: Normal, Mild, Moderate, Severe, and Proliferate_DR. Deep learning is one of the most popular methods for improving performance, particularly in the categorization and interpretation of medical images. Six deep-learning models—Custom CNN, Resnet50, Densenet121, EfficientNetB0, EfficientNetB2, and ViT—for the acceleration of diabetic retinopathy (DR) detection were evaluated using an extensive fundus image dataset that we got from Kaggle. With enhanced accuracy of 89%, precision of 89%, recall of 89%, and F1-score of 89% in a five-stage DR classification, the results show the superior performance of a DenseNet121 model.

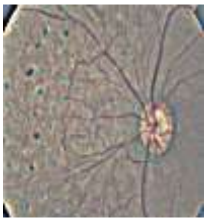
KEYWORDS— ***CNN, DensetNet121, DR, EfficientNetB0, EfficientNetB2, Resnet50, ViT***

1. INTRODUCTION

One of the most prevalent retinal diseases, DR is the main factor in human blindness (Mayo Clinic, 2018). It raises blood pressure in tiny blood vessels, which affects the retina's circulation and the light-sensitive tissue of the eye. Although several epidemiologic research and therapeutic trials have looked at the main risk factors for DR (such as hyperglycemia, hypertension, and dyslipidemia), there is still a great deal of diversity in the consistency, pattern, and severity of these risk variables (Nguyen, Quang H., Ramasamy Muthuraman, Laxman Singh, Gopa Sen, Anh Cuong Tran, Binh P. Nguyen, and Matthew Chua, 2020). There are five stages of DR (Nguyen, Quang H., Ramasamy Muthuraman, Laxman Singh, Gopa Sen, Anh Cuong Tran, Binh P. Nguyen, and Matthew Chua, 2020)i.e. No_DR, Mild, Moderate, Severe, Proliferate_DR as shown in Table I. The various signs and markers of diabetic retinopathy include microaneurysms, leaking blood vessels, retinal swellings, growth of abnormal new blood vessels, and damaged nerve tissues (Doshi, Darshit, Aniket Shenoy, Deep Sidhpura, and Prachi Gharpure, 2016) (Qummar, Sehrish, Fiaz Gul Khan, Sajid Shah, Ahmad Khan, Shahaboddin Shamshirband, Zia Ur Rehman, Iftikhar Ahmed Khan, and Waqas Jadoon.). In diabetic retinopathy, the absence of detectable signs characterizes the No Diabetic Retinopathy stage. Mild Diabetic Retinopathy presents with microaneurysms, and small swellings in retina blood vessels, while Moderate Diabetic Retinopathy is marked by the presence of hemorrhages and exudates in the retina. Severe Diabetic Retinopathy exhibits widespread hemorrhages, exudates, and signs of inadequate blood supply (ischemia) in the retina. Proliferative Diabetic Retinopathy (PDR) is distinguished by abnormal blood vessel growth (neovascularization) in the retina, posing risks such as retinal detachment, vitreous hemorrhage, and vision loss (Mayo Clinic, 2018) (Doshi, Darshit, Aniket Shenoy, Deep Sidhpura, and Prachi Gharpure, 2016). These stages represent a spectrum of retinal changes in diabetes as shown in Table I. There is a problem with manually diagnosing DR from retinal images because it is time-consuming and the number of specialists is less. To automate this procedure, a variety of deep learning techniques have been used, especially with datasets made available by Kaggle. Even with these improvements, current techniques still have difficulties in correctly differentiating between the different stages of DR, especially between Mild and Moderate DR and between Severe and Proliferative DR. Diabetic retinopathy (DR), also known as diabetic eye disease, is when damage occurs to the retina due to diabetes. It can eventually lead to blindness. It is an ocular manifestation of diabetes. Despite these intimidating statistics, research indicates that at least 90% of these new cases could be reduced if there were proper and vigilant treatment and monitoring of the eyes. The longer a person has diabetes, the higher his or her chances of developing diabetic retinopathy (Doshi, Darshit, Aniket Shenoy, Deep Sidhpura, and Prachi Gharpure, 2016) (Wan, Shaohua, Yan Liang, and Yin Zhang, 2018). In the healthcare field, the treatment of diseases is more effective when detected at an early stage. Diabetes is a disease that increases the amount of glucose in the blood caused by a lack of insulin (Gangwar, 2021) affects 425 million adults worldwide. Diabetes affects the retina, heart, nerves, and kidneys. The Nepal Diabetes Association reported that in urban areas diabetes affects approximately 15% of people aged 20 years and above. Diabetic eye care services in Nepal are not integrated with comprehensive diabetes management. Limited access to DR screening and vitreoretinal services are the major barriers to service utilization. As a result, people with diabetes often reach eye health providers with late-stage, sight-threatening DR (Qummar, Sehrish, Fiaz Gul Khan, Sajid Shah, Ahmad Khan, Shahaboddin Shamshirband, Zia Ur Rehman, Iftikhar Ahmed Khan, and Waqas Jadoon.). Diabetic Retinopathy (DR) is a complication of diabetes that causes the blood vessels of the retina to swell and to leak fluids and blood. DR can lead to a loss of vision if it is in an advanced stage. Worldwide, DR causes 2.6% of blindness (Hemanth, 2020) . The possibility of DR presence increases for diabetes patients who suffer from the disease for a long period (Carrera, Enrique V., Andrés González, and Ricardo Carrera) (Khan, Zubair, Fiaz Gul Khan, Ahmad Khan, Zia Ur Rehman, Sajid Shah, Sehrish Qummar, Farman Ali, and Sangheon Pack.). To prevent blindness, diabetics should be screened every year. There is increasing progress in collaboration between a diabetes care physician and an ophthalmologist. A common practice for detecting diabetic eye disease is to examine the fundus image and assess the severity of the disease (Dipesh Gyawali, Alok Regmi, Aatish Shakya, Ashish Gautam, Surendra Shrestha, 2020). There are a few opportunities for diabetic patients to undergo fundus examinations in areas with few doctors. Further, there is a shortage of ophthalmologists who can diagnose and treat diabetic retinopathy (Dipesh Gyawali, Alok Regmi, Aatish Shakya, Ashish Gautam, Surendra Shrestha, 2020). The current clinical diagnosis of diabetic retinopathy (DR) heavily relies on the manual examination of colour fundus images by ophthalmologists, which is time-consuming, prone to errors, and highly dependent on the expertise of the clinician. This process leads to delayed detection and treatment, particularly in areas with limited medical resources, resulting in irreversible visual loss and even blindness for many patients. Despite previous efforts in using image classification and machine learning techniques for DR detection, the lack of high-tech medical facilities in many regions exacerbates the challenge of early diagnosis. Therefore, there is a critical need to develop an automated DR detection system that can operate effectively in resource-constrained settings, enabling early identification and intervention for patients at risk of visual impairment. The objectives of the proposed method is to develop a methodology that can accurately detect five stages of diabetes using deep learning and compare deep learning models. The model can be trained on a larger dataset to improve the accuracy of the classification. Explainability AI can be integrated for the explanation of certain regions responsible for the specific stage.

Table 1. Retina Images for five DR stages and their features.

Stages	Name	Explanation	Sample image
1	No_DR	No Diabetes	
2	Mild_DR	The earliest stage where only microaneurysms can happen	
3	Moderate_DR	The ability of blood transportation due to their distortion and swelling with the progress of the disease	
4	Severe_DR	Blood supply to the retina due to increase blockage of more blood vessels	

5	PDR	The advance stage where the growth feature secreted by the retina activate proliferation of the new blood vessel	

1.1 Literature Reviews

Different research work has been performed on the classification of DR stages. In the given section, recent related work and their proposed methodologies are explained. Researchers have proposed different models for classifying different DR stages; this section discusses the most prominent techniques. Wan, Shaohua, et al.perform diabetic retinopathy detection by image classification using deep learning and attempt to classify fundus images using transfer learning and CNN on DR images and get better accuracy of 95.68% (Kaza, Silpa Yao, Lisa C. Bhada-Tata, Perinaz Van Woerden, Frank, 2018). Nguyen, Quang H., et al. present an automated classification system of DR screening using ML models such as CNN, VGG-16, and VGG-19 and achieves 80% sensitivity, 82%accuracy 82%specificity for classifying images into 5 categories (Nguyen, Quang H., Ramasamy Muthuraman, Laxman Singh, Gopa Sen, Anh Cuong Tran, Binh P. Nguyen, and Matthew Chua, 2020). Gangwar, Akhilesh Kumar, and Vadlamani Ravi performed Diabetic retinopathy detection using transfer learning and deep learning making a hybrid model of CNN on top of Inception-ResNet-v2 and evaluated the performance by achieving a test accuracy of 72.33% and 82.18% on different dataset (Dipesh Gyawali, Alok Regmi, Aatish Shakya, Ashish Gautam, Surendra Shrestha, 2020). S. Qummar *et al* ‘s A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection, trains an ensemble of five deep Convolution Neural Network (CNN) models (Resnet50, Inceptionv3, Xception, Dense121, Dense169) to encode the rich features and improve the classification for different stages of DR (Hannan, M.A., Arebey, Maher, Begum, R.A., and Hassan Basri, 2011). D. Doshi. et al. paper on “Diabetic retinopathy detection using deep convolutional neural networks" aims at automatic diagnosis of the disease in its different stages by implementing GPU accelerated deep convolutional neural networks to automatically diagnose and gaining the single model accuracy of the CNN 0.386 on a quadratic weighted kappa metric and ensembled of three such similar models resulted in a score of 0.3996 (Doshi, Darshit, Aniket Shenoy, Deep Sidhpura, and Prachi Gharpure, 2016). Z. Khan et al performed Diabetic Retinopathy Detection Using VGG-NIN where VGG-NIN can process a DR image at any scale due to the SPP layer's virtue and better classification due to the stacking of NiN adds extra nonlinearity to the model. The experimental results show that the proposed model performs better in terms of accuracy, and computational resource utilization compared to state-of-the-art methods (Karthik, M., Sreevidya, L., Nithya Devi, R., Thangaraj, M., Hemalatha, G., and R. Yamini., 2023). Hemanth, D. Jude. et al proposed the employment of image processing by the classification of a CNN with histogram equalization, and the contrast limited adaptive histogram equalization and was validated using 400 retinal fundus images within the MESSIDOR database, and average values for different performance evaluation parameters were obtained as accuracy 97%, sensitivity (recall) 94%, specificity 98%, precision 94%, FScore 94%, and G-Mean 95% (Sudha, S., Vidhyalakshmi, M., Pavithra, K, 2016).E. V. Carrera. et al. proposed a computer-assisted diagnosis to automatically classify the grade of non-proliferative diabetic retinopathy at any retinal image using a support vector machine to figure out the retinopathy grade of each retinal image obtaining a maximum sensitivity of 95% and a predictive capacity of 94% (Mittal, G., Yagnik, K.B., Garg, M., Krishnan, N.C., 2016).

2. COMPUTATIONAL AND THEORETICAL DETAILS

2.1 Custom CNN Architecture

A convolutional neural network (CNN) (A. Krizhevsky, 2012) is a type of artificial neural network used primarily for image recognition and processing, due to its ability to recognize patterns in images. In the custom CNN architecture as shown in Figure 1, each convolutional layer employs 3x3 kernels with a stride of 1 and Rectified Linear Unit (ReLU) activation functions, facilitating feature extraction and non-linearity. Additionally, max-pooling layers with 2x2 windows and a stride of 2 are utilized to downsample feature maps effectively. The number of filters in each of the four convolutional layers is gradually increased from 128 to 1024, enabling the network to capture increasingly complex features as the depth of the network increases. Dropout layers with a dropout rate of 0.45 are strategically inserted after each fully connected layer to mitigate overfitting during the training process. Moreover, batch normalization is applied with a momentum of 0.99 and an epsilon of 0.001 along axis -1, contributing to the robustness of the model by normalizing the activations of each layer. To optimize the model, the Adamax optimizer is employed with a learning rate of 0.001, while the categorical cross-entropy loss function is utilized to quantify the disparity between predicted and actual class labels, facilitating effective training and convergence.

2.2 Resnet 50

The ResNet50 (P, 2023) architecture, proposed by He et al. in their paper "Deep Residual Learning for Image Recognition," serves as the backbone of our image classification system. ResNet50 as shown in

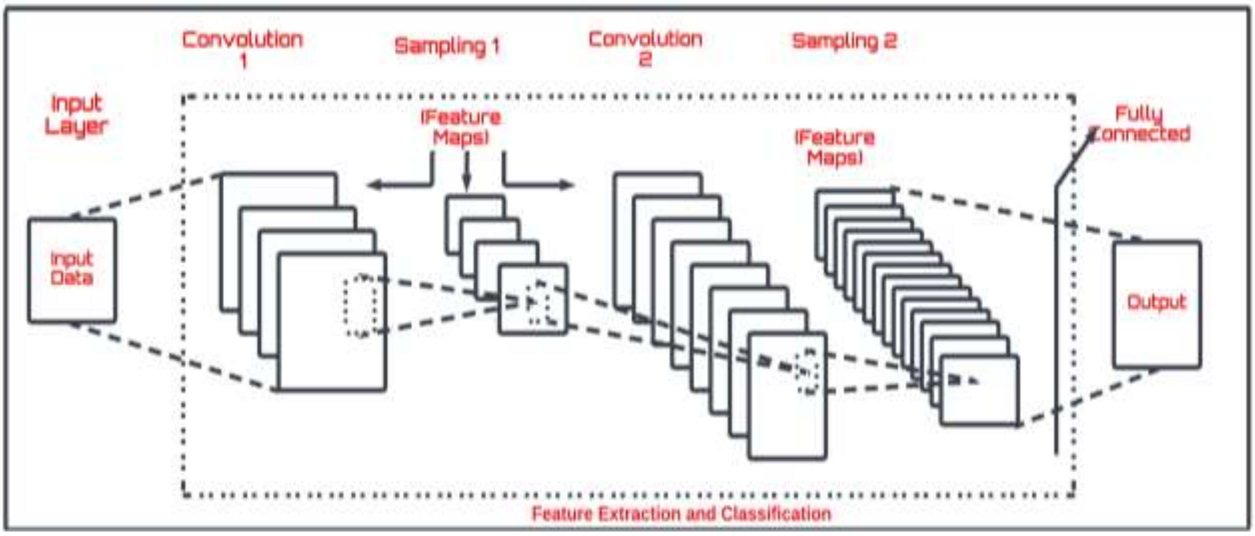


Figure 1. Custom CNN Architecture

Figure 2 is constructed using a series of convolutional layers, residual blocks, and global average pooling layers. In our implementation, we utilized the pre-trained ResNet50 model available through the TensorFlow Keras API with weights initialized from the ImageNet dataset. By leveraging pre-trained weights, the model benefits from learned feature representations, facilitating robust performance even with limited training data. We modified the top layers of the ResNet50 model to suit our specific classification task, adding batch normalization, dropout, and fully connected layers with appropriate activation functions to ensure model generalization and prevent overfitting. The basic idea of this shortcut connection is shown in Figure 2.

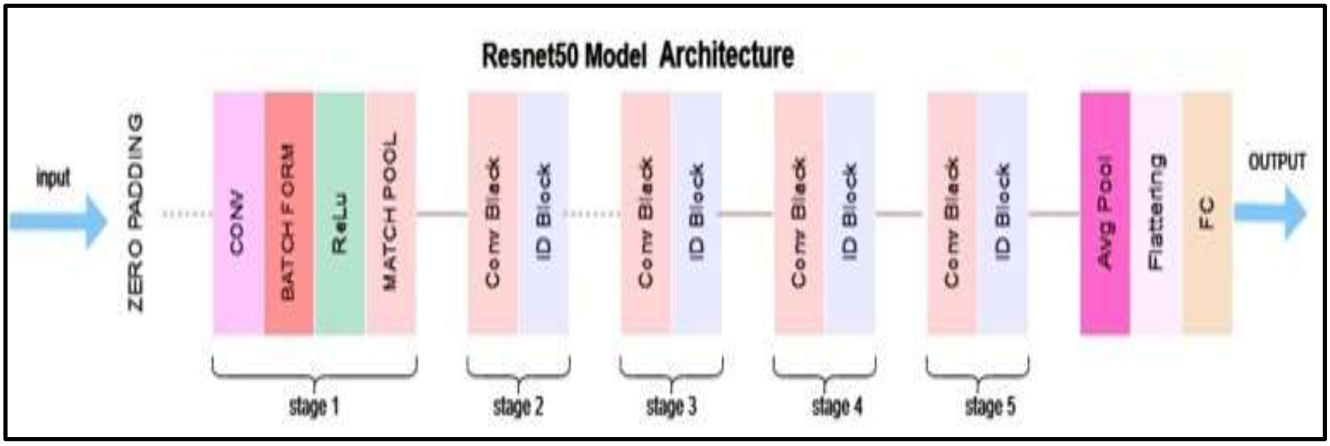


Figure 2. A Resnet 50 basic block

2.3 DenseNet121

The DenseNet121 (Yang M. T., 2016) architecture as shown in Figure 3, introduced by Huang et al. in their paper "Densely Connected Convolutional Networks," forms the core of our image classification framework. DenseNet121 is characterized by dense connectivity patterns, wherein each layer receives feature maps from all preceding layers, promoting feature reuse and facilitating gradient flow throughout the network. Constructed with dense blocks comprising convolutional layers, batch normalization, and Rectified Linear Unit (ReLU) activations, DenseNet121 encourages feature aggregation and fosters deep feature representations. In our implementation, we leveraged the pre-trained DenseNet121 model available through the TensorFlow Keras API, initialized with weights obtained from the ImageNet dataset. Using pre-trained weights enables the model to harness learned feature representations, enhancing its performance even when trained on limited data. To adapt DenseNet121 to our specific classification task, we customized the top layers of the model, integrating batch normalization, dropout, and fully connected layers with suitable activation functions. This customization facilitates model generalization and aids in mitigating overfitting concerns. By adopting DenseNet121 as the backbone of our image classification system and tailoring it to our specific requirements, we ensure robust performance and effective utilization of available resources.

2.4 EfficientNetB0 and EfficientNetB2 Architecture

The EfficientNetB0 and EfficientNetB2 architectures as shown in Figure 4 (Adedeji, 2019), pioneered by Tan et al. in their paper "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," form the cornerstone of our image classification framework. These architectures are crafted to deliver superior performance while maintaining computational efficiency, making them ideal for various computer vision tasks. EfficientNetB0 serves as the baseline model in the EfficientNet series, featuring a hierarchical structure of convolutional layers, where each layer is meticulously optimized for efficiency. The hallmark of EfficientNet lies in its compound scaling strategy, which uniformly scales the network's depth, width, and resolution to achieve an optimal balance between model size and performance. In our implementation, we harnessed the pre-trained EfficientNetB0 and EfficientNetB2 models provided by the TensorFlow Keras API, initialized with weights derived from the ImageNet dataset. Leveraging these pre-trained weights enables the model to capitalize on learned feature representations, enhancing performance even with limited training data. Additionally, we tailored the top layers of the EfficientNetB0 and EfficientNetB2 model to our specific classification task, incorporating techniques such as batch normalization, dropout, and fully connected layers with suitable activation functions to foster model generalization and mitigate overfitting concerns.

2.5 Vision Transformer (ViT)

In our methodology, we employed a pre-trained Vision Transformer (ViT) architecture as shown in Figure 5 (Pathak, 2017), a novel approach in computer vision introduced by Dosovitskiy et al., which replaces traditional convolutional layers with self-attention mechanisms. Unlike traditional convolutional neural networks (CNNs), ViT treats images as sequences of patches and applies self-attention mechanisms to capture global dependencies among them. This enables ViT to effectively model long-range interactions within images, facilitating a better understanding of spatial relationships and semantic contexts. Specifically, we utilized a ViT model with a patch size of 16x16, implemented in PyTorch. In addition to its self-attention mechanisms, the ViT architecture utilizes **a** transformer encoder to process input image patches, enabling it to capture long-range dependencies effectively. By replacing convolutions with self-attention, ViT achieves competitive performance on various computer vision tasks while offering greater flexibility and scalability. The ViT model was fine-tuned on our dataset for 50 epochs, leveraging techniques such as data augmentation and learning rate scheduling to enhance performance. Through the Hugging Face Transformers library, we seamlessly integrated the pre-trained ViT model into our workflow, enabling efficient training and evaluation. The trained ViT model was evaluated using standard metrics such as accuracy and confusion matrices to assess its performance in image classification tasks.

2.6 process model

As shown in Figure 6 the process model input image i.e. fundus image is sent as an input. Due to data imbalance, data up-sampling is done to balance the datasets across five classes. Then image preprocessing is done. In image processing, we have done image cleaning where the image that doesn’t satisfy the specific classes has been deleted from datasets. For example, if the mild images are in the moderate images class, then it is removed manually. Then all the images are resized to a fixed size (224,224) then various augmentation techniques are applied like horizontal flipping, width shift, height shift, and zoom. Then the datasets are passed through the various deep learning algorithms like custom CNN, ResNet50, DenseNet121, EfficientNetB0, EfficientNetB2 and ViT. Then finally the output is taken out based on the algorithm for specific input. Finally, the evaluation is done for various algorithms using evaluation criteria like accuracy, confusion matrix, precision, recall etc.

2.7 Activity Diagram

The activity diagram as shown in Figure 7 outlines the sequential flow of actions within the diabetic retinopathy detection system. It illustrates the steps involved in the detection process, including image acquisition, preprocessing, feature extraction, classification, and result presentation. This diagram aids in understanding the overall workflow of the system. We follow the following procedure as in Figure 7 with the help of CNN: Data collection and preprocessing: The first step in creating a prototype model is to collect data on the different types of fundus images of the eye that need to be classified. Once the data is collected, it needs to be pre-processed to create a clean and uniform dataset that can be used for training the model. This can include resizing the images, removing noise, and normalizing the data.

Model architecture: The next step is to design the architecture of the prototype model. This involves selecting the appropriate number of layers, activation functions, and other hyperparameters that will be used in the model. The goal is to create a model that can accurately classify the stages of diabetes based on their features.

Training: After designing the model architecture, the next step is to train the prototype model using the pre-processed data. The training process involves adjusting the weights and biases of the model to minimize the difference between the predicted and actual outputs. This is done using a training algorithm that adjusts the weights and biases based on the error between the predicted and actual outputs.

Validation: Once the prototype model is trained, it must be validated to ensure it is not overfitting the training data. This involves testing the model on a separate set of data that was not used in the training process. If the model performs well on the validation data, it can be concluded that the model is generalizing well and is not overfitting. **Results:** Finally, the accuracy and other performance metrics obtained from the prototype model can be reported. This can include metrics such as precision, recall, and F1 score, as well as the confusion matrix. The report can also include any insights gained from the testing and validation process.

3. RESULTS AND DISCUSSIONS

The diabetic retinopathy detection using deep learning algorithms was implemented and evaluated. The results obtained from the experiments are presented in this section.

Model Training:

Our study on diabetic retinopathy detection utilized various transfer learning algorithms, including ResNet50, DenseNet121, EfficientNetB0, EfficientNetB2, custom CNN, and Vision Transformer (ViT). While most algorithms were implemented in TensorFlow and trained for 20 epochs, ViT was implemented in PyTorch and trained for 50 epochs. Despite using ViT, which represents a state-of-the-art approach, it did not yield superior results compared to DenseNet121. The model was trained on 80% of the training dataset,10%validation dataset, and 10%testing data set. The training process was conducted over 20 epochs, with a batch size of 40, using the Adamax optimizer and the categorical cross-entropy loss function. Different model has different accuracy Custom CNN has an accuracy has 55% Resnet50 has an accuracy of 83% DenseNet121 has an accuracy of 89 %, EfficientNetB0 has an accuracy of 84%, EfficientNetB2 has an accuracy of 87%, ViT has an accuracy of 71%. The training process prioritized achieving a training accuracy exceeding 90%, after which validation loss was monitored. Additionally, the learning rate was adjusted after a patience period of 5 epochs—this adaptive approach aimed to balance model accuracy with generalization performance, ensuring robustness in diabetic retinopathy detection

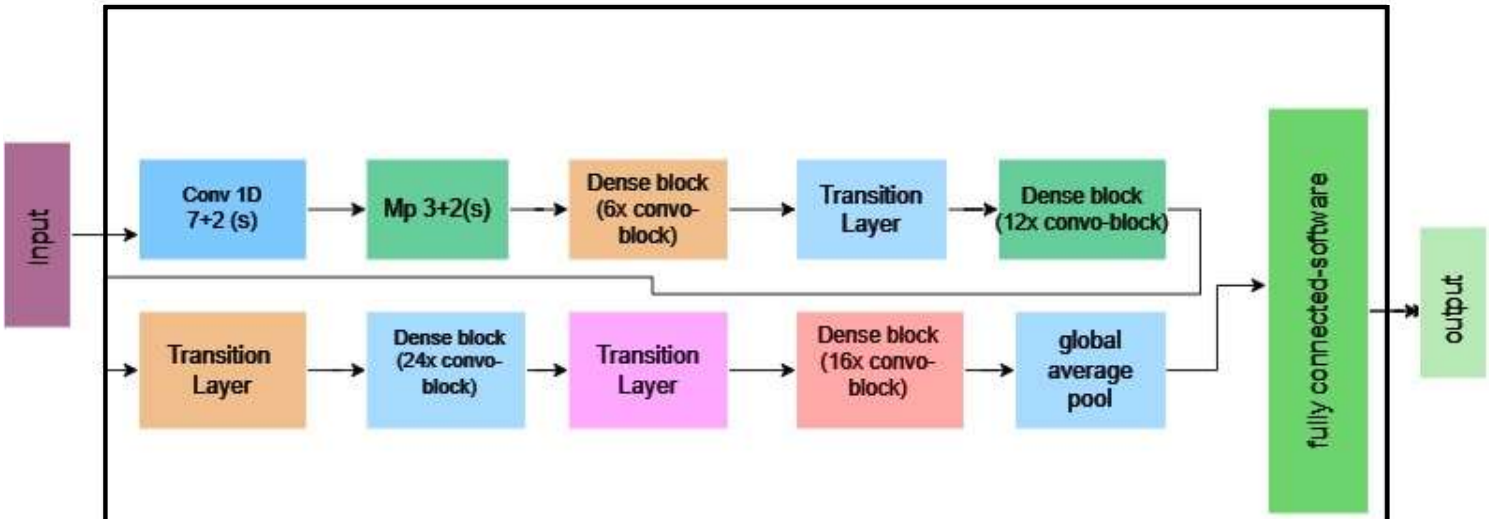


Figure 3. A Densenet121 basic block

3.1 Visualization of Accuracy and Loss

The accuracy and loss curves for the training and validation sets are shown in Figures 8-13 The left graph shows the training accuracy vs validation accuracy over epochs and the right graph shows training loss vs validation loss over epochs. From all the accuracy and loss graphs it can be seen that the model is performing well on training datasets. Overall, from the visualization graph, densenet121 validation accuracy is high and validation loss is low.The graph is shown in Figurew 8-13 depicts training and validation

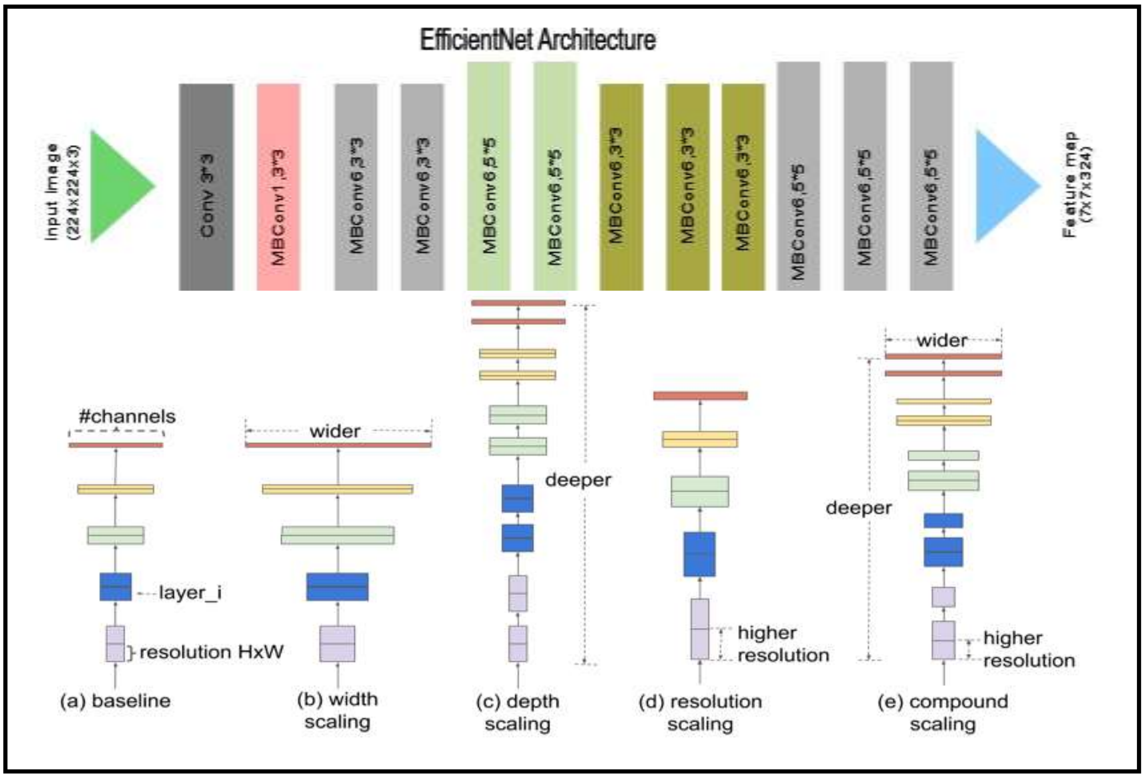


Figure 4. An EfficientnetB0 and efficientnetB2 basic blocks

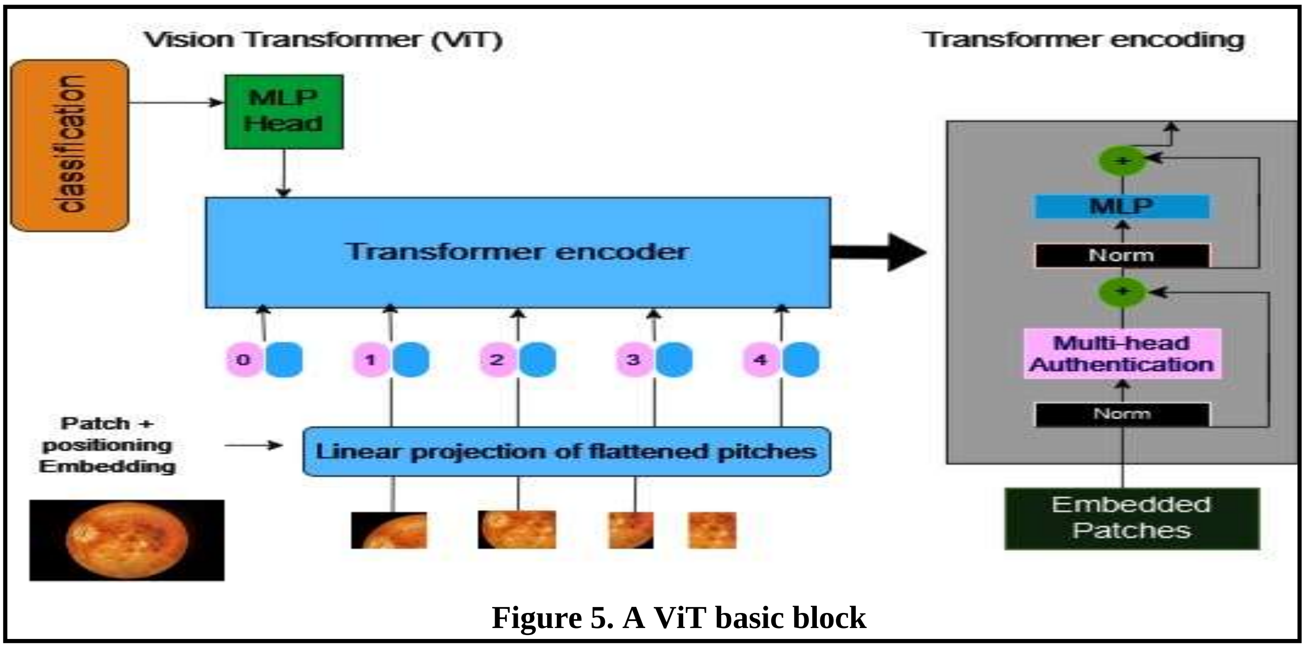


Figure 5. A ViT basic block

accuracy (blue and orange lines) and loss (blue and orange lines) over epochs as in Figure (a) and Figure (b) respectively. Training accuracy steadily rises, indicating pattern recognition improvement, while validation accuracy increases more slowly,

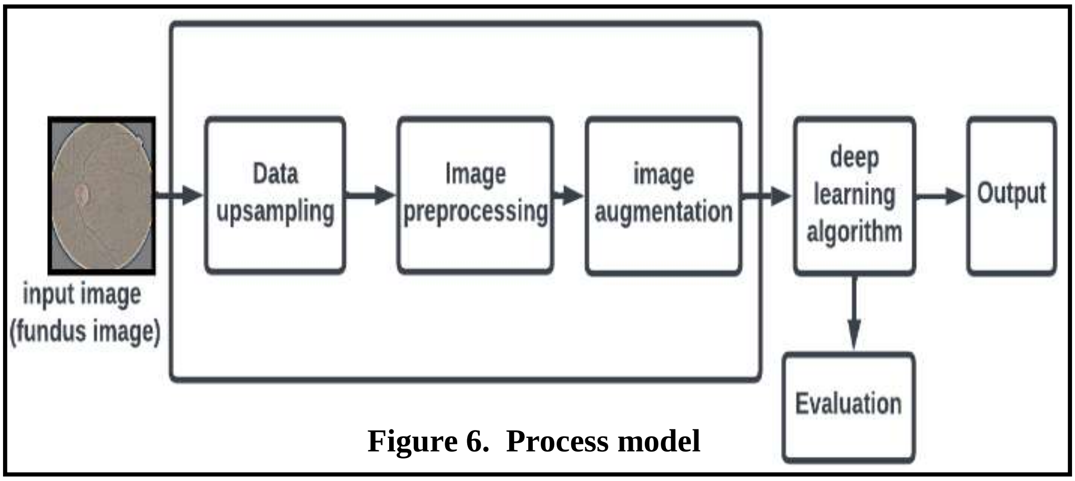


Figure 6. Process model

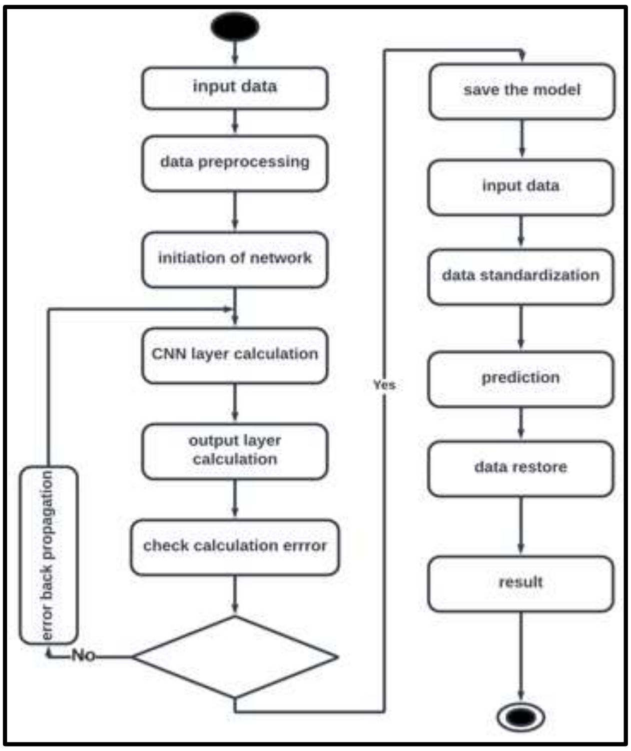


Figure 7. Activity Diagram

suggesting potential overfitting. Training loss decreases as epochs progress, signifying better prediction accuracy, though validation loss decreases at a slower pace, hinting at overfitting concerns. This visualization underscores the balance between model performance and generalization during training. The potential overfitting may be due to the complexity of the model, insufficient data, hyperparameter tuning, early stopping, and data quality. The model might still be too complex for the given dataset, allowing it to capture noise in the training data rather than meaningful patterns. Despite data augmentation, the dataset may still be relatively small, making it challenging for the model to generalize effectively. Inadequate tuning of hyperparameters such as learning rate, batch size, or regularization strength can also contribute to overfitting. If there are inconsistencies or inaccuracies in the training data, the model may learn from these errors, leading to overfitting. In such cases, reducing model complexity or employing simpler architectures could help mitigate overfitting. Gathering more diverse and representative data could help address insufficient data issues. Including domain related i.e. medical preprocessing techniques may address this issue. Continuously optimizing hyperparameter parameters through experimentation may reduce the potential risk of overfitting. Ensuring high-quality data may reduce the potential risks. By carefully considering these factors and continually refining the model and training

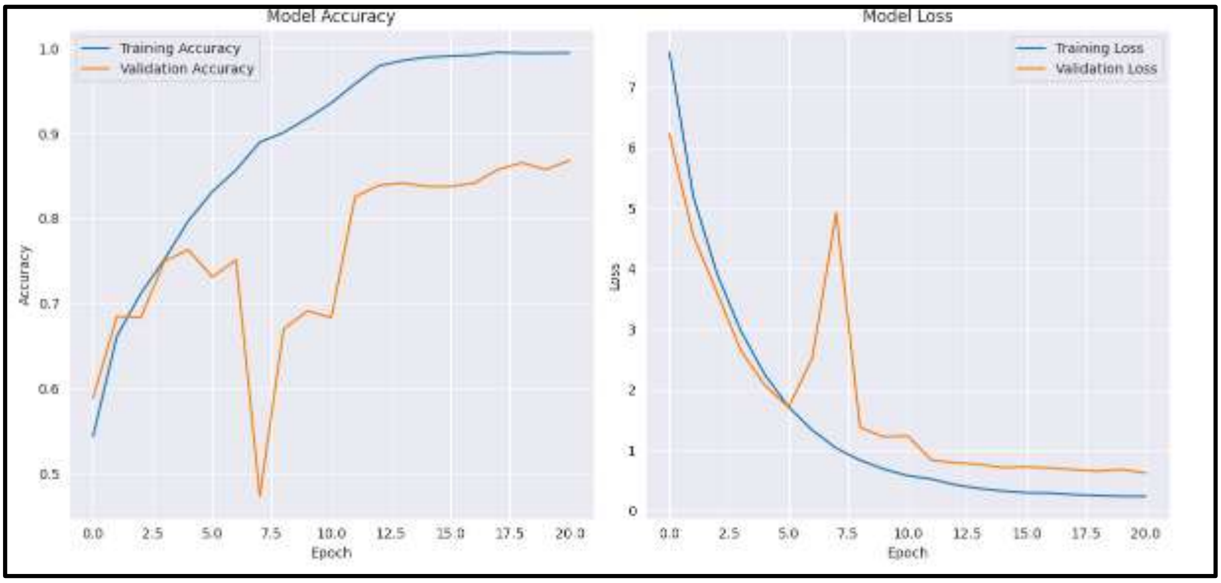


Figure 8. (a)Accuracy vs Epoch; (b) Loss vs Epoch of ResNet50

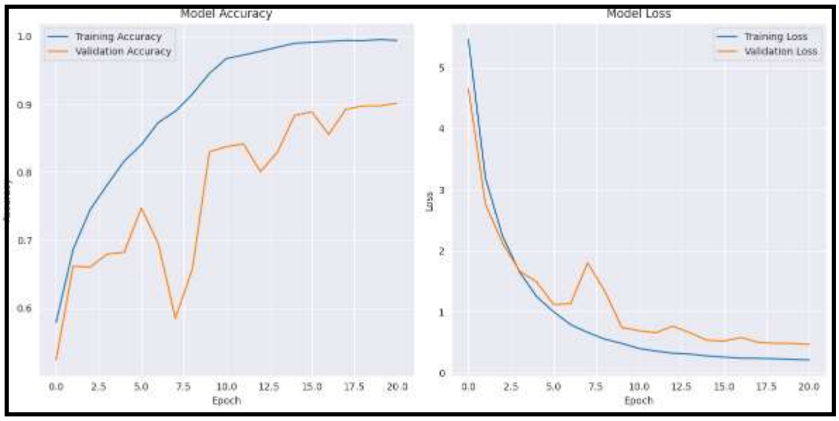


Figure 9. (a) Accuracy vs Epoch (b) Loss vs Epoch of DenseNet121

process, the potential for overfitting can be minimized, ensuring the model generalizes well to the

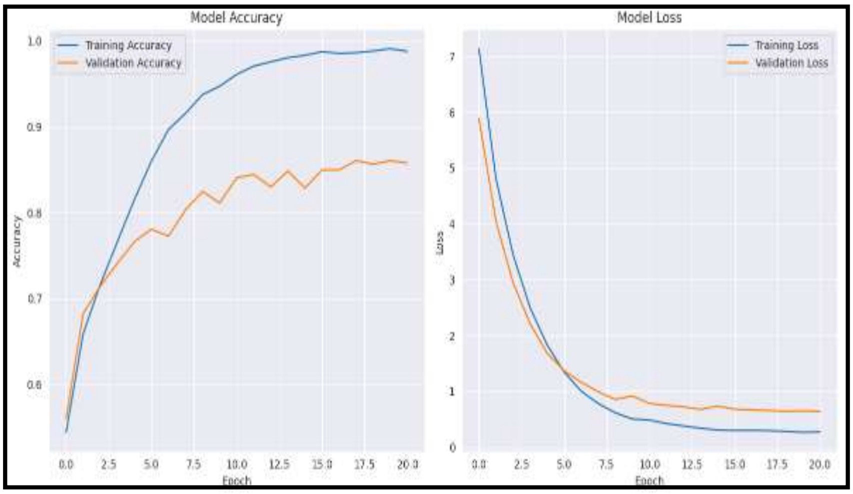


Figure 10. (a)Accuracy vs Epoch (b) Loss vs Epoch of EfficientNetB0

unseen data.

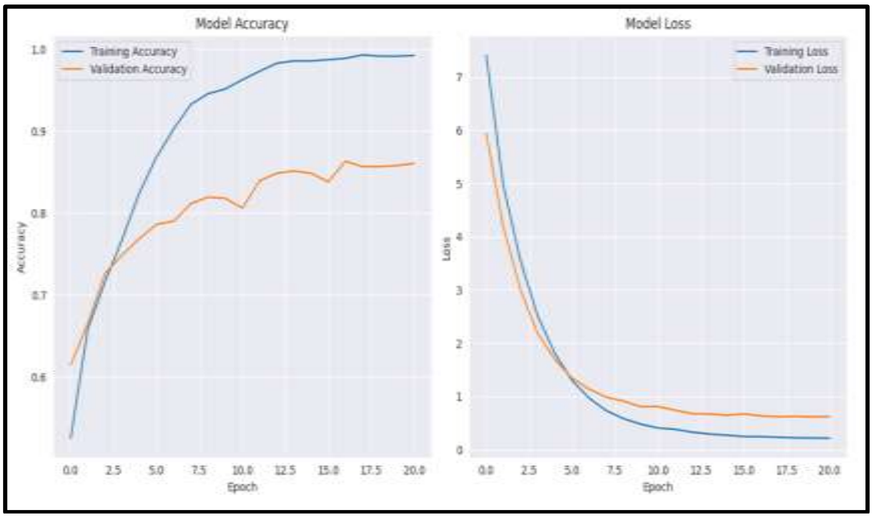


Figure 11. (a)Accuracy vs Epoch (b) Loss vs Epoch of EfficientNetB2

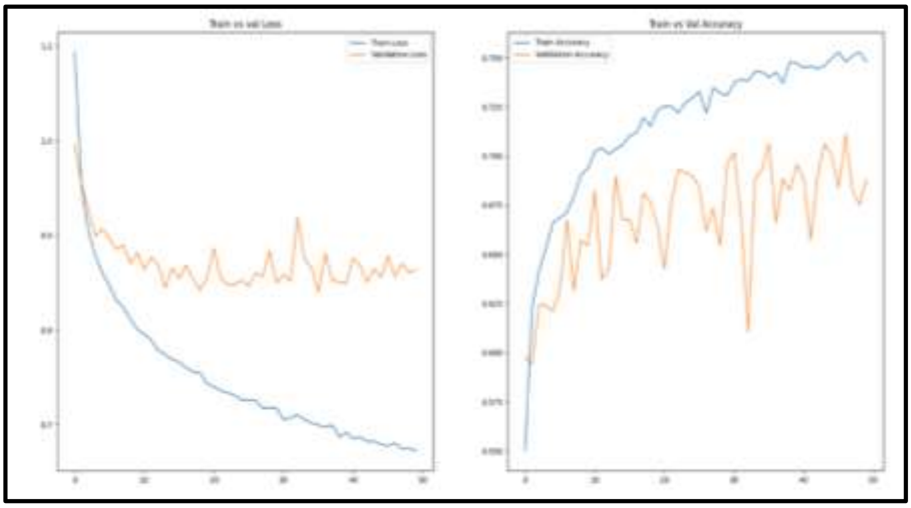


Figure 12. (a)Accuracy vs Epoch (b) Loss vs Epoch of ViT

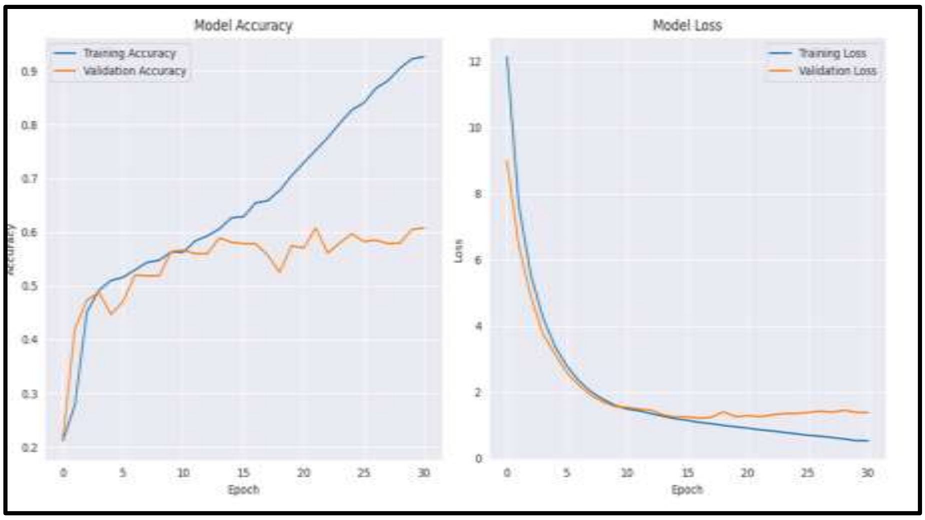


Figure 13. (a)Accuracy vs Epoch (b) Loss vs Epoch of Custom

3.2 Classification Report Analysis

Table 1 Performance measure of each class on Precision

Labels	Custom CNN	Resnet50	DenseNet121	EfficientNet-B0	EfficientNet-B2	ViT
No_DR	0.51	0.95	0.98	0.94	0.97	1.0
Mild	0.61	0.82	0.85	0.81	0.86	0.62
Moderate	0.92	0.71	0.89	0.77	0.76	0.38
Severe	0.33	0.82	0.84	0.82	0.91	0.62
Proliferate_DR	0.44	0.85	0.86	0.87	0.82	0.73

Table 2 Performance measure of each class on Recall

Labels	Custom CNN	Resnet50	DenseNet121	EfficientNet-B0	EfficientNet-B2	ViT
No_DR	0.34	0.99	0.99	0.99	0.97	1.0
Mild	0.50	0.83	0.88	0.80	0.88	0.75
Moderate	0.99	0.68	0.75	0.70	0.71	0.46
Severe	0.36	0.89	0.89	0.93	0.91	0.55
Proliferate_DR	0.61	0.77	0.92	0.80	0.85	0.61

Table 3 Performance measure of each class onF1-Score

Labels	Custom CNN	Resnet50	DenseNet121	EfficientNet-B0	EfficientNet-B2	ViT
No_DR	0.41	0.97	0.98	0.97	0.97	1.0
Mild	0.55	0.82	0.87	0.81	0.87	0.68
Moderate	0.95	0.70	0.81	0.73	0.73	0.42
Severe	0.35	0.85	0.87	0.87	0.91	0.58
Proliferate_DR	0.51	0.81	0.89	0.84	0.83	0.66

The performance measures across different classes for each model, as presented in Table 2, Table 3, and Table 4, reveal notable variations in precision, recall, and F1 scores. For instance, while certain models like Resnet50 and DenseNet121 exhibit consistently high precision across most classes, Custom CNN demonstrates comparatively lower precision in some classes. However, in terms of recall, Custom CNN often lags behind other models, particularly in classes such as No_DR and Severe. This indicates a trade-off between precision and recall across models, where some excel in one metric but fall short in the other. Additionally, the F1-score, which balances both precision and recall, further highlights the performance differences, showcasing the overall effectiveness of each model in capturing true positives while minimizing false positives and false negatives. These observations underscore the importance of considering multiple performance metrics to evaluate the performance of classification models across various classes comprehensively.

3.3 Analysis using Confusion Matrix

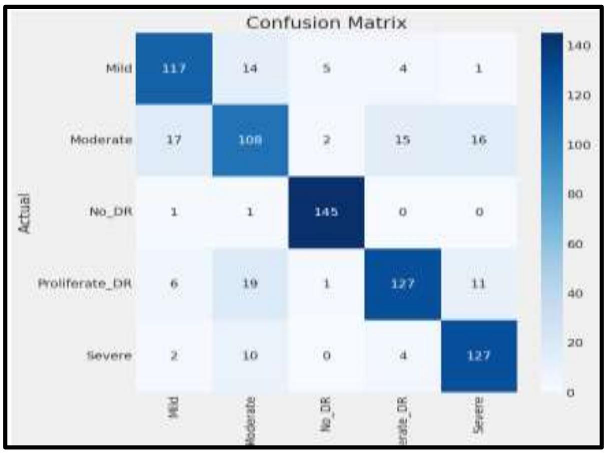


Figure 14. Confusion matrix of ResNet50

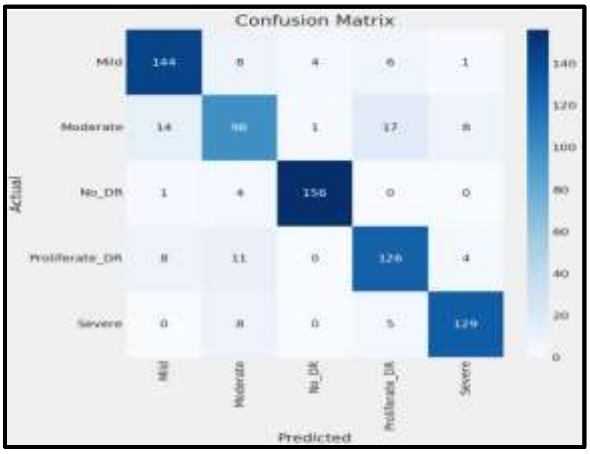


Figure 15. Confusion matrix of DenseNet121

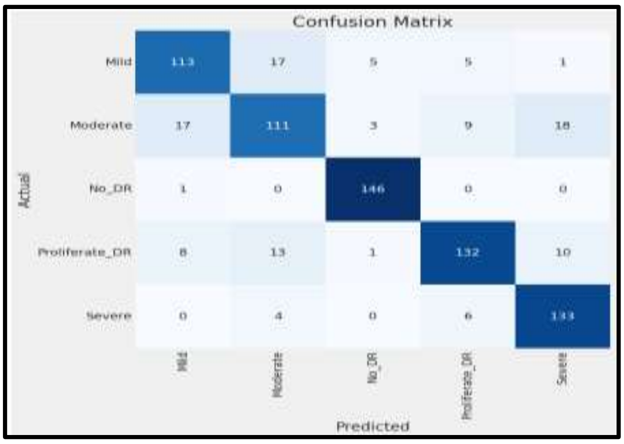


Figure 16. Confusion matrix of EfficientNetB0

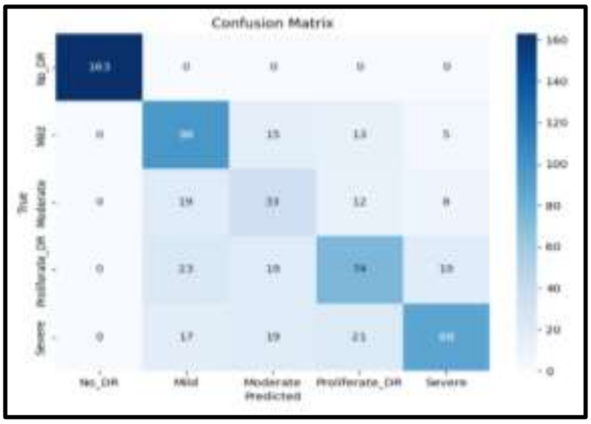


Figure 17. Confusion matrix of EfficientNetB2

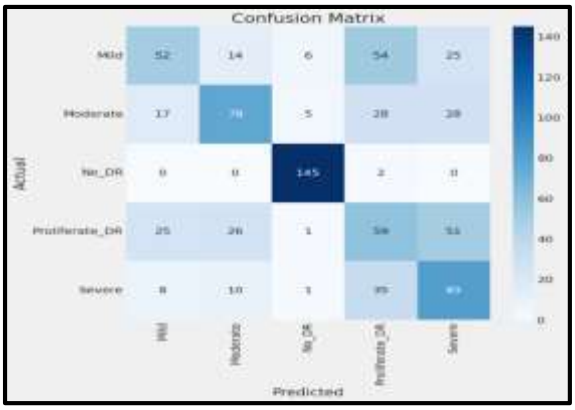


Figure 18. Confusion matrix of ViT

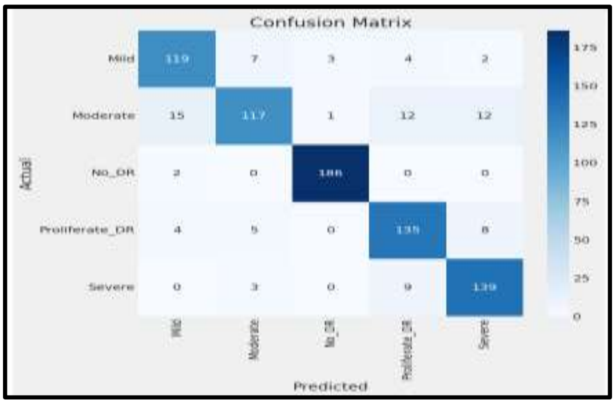


Figure 19. Confusion matrix of Custom CNN

The following things are observed from the above confusion matrices:

(1) ViT as shown in Figure 18 predicts all No_DR (no diabetic retinopathy) images as No_DR, indicating potential bias towards the majority class. This suggests a need for further optimization or class-balancing techniques to improve performance in minority classes.

(2) All algorithms perform well in predicting most No_DR images correctly, reflecting their ability to handle the majority class effectively. This observation underscores the importance of addressing class imbalance issues and mitigating model bias.

(3) Custom CNN struggles to correctly identify mild and proliferate diabetic retinopathy (DR) images. Potential reasons for this difficulty may be architectural limitations as shown in Figure 19.

(4) All algorithms face challenges in accurately identifying mild, moderate, and proliferate DR images. This highlights the complexity of distinguishing between different severity levels of DR and the need for advanced modeling techniques. Also, Figures 14, 16, and 17 display confusion matrices for the classification performance of several models. In particular, the confusion matrix of ResNet50 is shown in Figure 14, that of EfficientNetB0 is shown in Figure 16, and that of EfficientNetB2 is shown in Figure 17.

(5) DenseNet121 as shown in Figure 15 demonstrates superior performance compared to other algorithms in identifying classes. Its architecture, featuring dense connectivity and feature reuse mechanisms, contributes to its effectiveness in complex classification tasks like DR identification.

4. Conclusions

In this investigation of six algorithms for diabetic retinopathy detection of 5 stages - ResNet50, DenseNet121, EfficientNetB0, EfficientNetB2, custom CNN, and Vision Transformer (ViT)—DenseNet121 emerges as the standout performer, exhibiting high accuracy in classifying retinal images indicative of diabetic retinopathy. This underscores the efficacy of DenseNet121 in leveraging pre-trained features to accurately identify pathological changes in retinal structures associated with the condition. While DenseNet121 showcased superior performance, the study also highlights the competitive capabilities of the other models, including ResNet50, EfficientNetB0, EfficientNetB2, custom CNN, and Vision Transformer (ViT), indicating the versatility of transfer learning methodologies in medical image analysis. However, Vision Transformer (ViT) did not demonstrate satisfactory accuracy in this study, indicating potential limitations in its applicability for diabetic retinopathy detection. Future research endeavors may focus on refining model generalization through the incorporation of diverse datasets and the real world, ultimately advancing the accuracy and efficiency of diabetic retinopathy detection for improved patient outcomes.

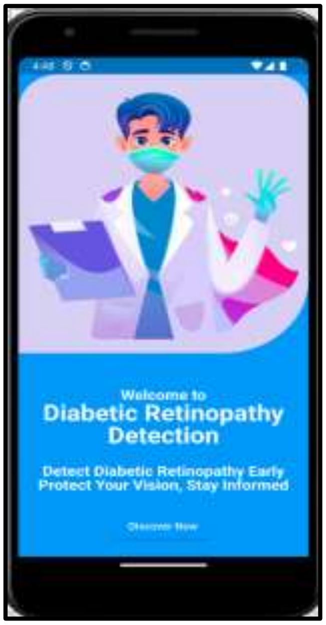
5. Limitations and Future Work

There are some issues with the current models for detecting diabetic retinopathy, including picture quality differences, trouble identifying advanced stages, biases related to demographics, and poor interpretability. To improve management and provide more individualized care, future research should concentrate on expanding the capacity to anticipate the course of the disease and other ocular disorders, utilizing interpretable methodologies, integrating multimodal data, expanding the use of domain adaptation and data augmentation, and diversifying datasets.

Acknowledgment

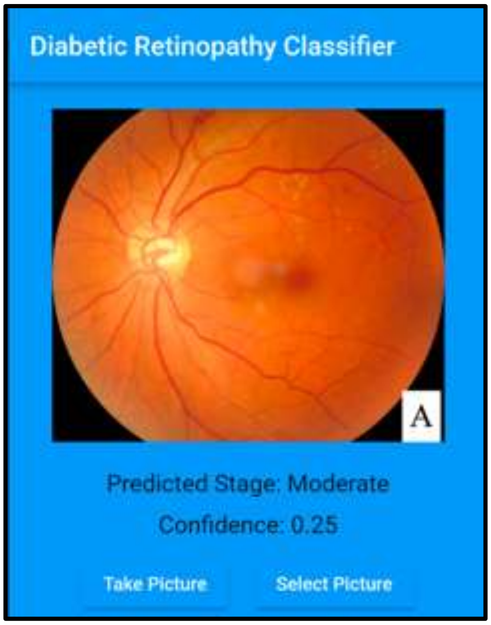
We would like to express our profound appreciation to Department of Computer and Electronics, Communication, and Information Engineering, Kathford International College of Engineering and Management for the steadfast assistance throughout this study. We especially appreciate Er. Jalauddin Mansur, the project coordinator for his invaluable advice and support. Finally, we acknowledge the entire editorial and reviewer board members (KJEM) for the constructive reviews.

SUPPLIMENTARY MATERIALS (IMAGES)



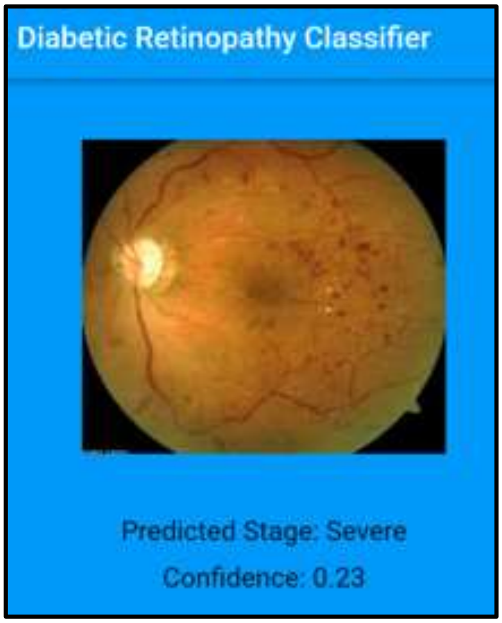
Welcome to
Diabetic Retinopathy Detection

Detect Diabetic Retinopathy Early
Protect Your Vision, Stay Informed



Diabetic Retinopathy Classifier

Predicted Stage: Moderate
Confidence: 0.25



Diabetic Retinopathy Classifier

Predicted Stage: Severe
Confidence: 0.23

REFERENCES

1. *Mayo Clinic*. (2018). (Diabetic Retinopathy) Available:https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611

2. Nguyen, Quang H., Ramasamy Muthuraman, Laxman Singh, Gopa Sen, Anh Cuong Tran, Binh P. Nguyen, and Matthew Chua. (2020). Diabetic retinopathy detection using deep learning. *Proceedings of the 4th international conference on machine learning and soft computing*, (pp. 103-107).

3. Doshi, Darshit, Aniket Shenoy, Deep Sidhpura, and Prachi Gharpure. (2016). Diabetic retinopathy detection using deep convolutional neural networks. *International conference on computing, analytics and security trends (CAST)* (pp. 261-266). IEEE.

4. Wan, Shaohua, Yan Liang, and Yin Zhang. (2018). Deep convolutional neural networks for diabetic retinopathy detection by image classification. *Computers & Electrical Engineering*, *72*, 274-282.
5. Gangwar, A. K. (2021). Diabetic retinopathy detection using transfer learning and deep learning. In S. Singapore (Ed.), *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020)*, *1*, 679-689.

6. Qummar, Sehrish, Fiaz Gul Khan, Sajid Shah, Ahmad Khan, Shahaboddin Shamshirband, Zia Ur Rehman, Iftikhar Ahmed Khan, and Waqas Jadoon. (n.d.). A deep learning ensemble approach for diabetic retinopathy detection. *7*, 150530 - 150539.

7. Khan, Zubair, Fiaz Gul Khan, Ahmad Khan, Zia Ur Rehman, Sajid Shah, Sehrish Qummar, Farman Ali, and Sangheon Pack. Diabetic retinopathy detection using VGG-NIN a deep learning architecture. 61408-61416.

8. Hemanth, D. J. (2020). An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network. *Neural Computing and Applications*, *32*, 707-721.

9. Carrera, Enrique V., Andrés González, and Ricardo Carrera (2017). Automated detection of diabetic retinopathy using SVM. *IEEE XXIV international conference on electronics, electrical engineering and computing (INTERCON)* (pp. 1-4).

10. Dipesh Gyawali, Alok Regmi, Aatish Shakya, Ashish Gautam, Surendra Shrestha. (2020). Comparative Analysis of Multiple Deep CNN Models for Waste Classification. *5th International Conference on Advanced Engineering and ICT-Convergence*, *1*, 6.

11. Kaza, Silpa Yao, Lisa C. Bhada-Tata, Perinaz Van Woerden, Frank. (2018). *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050.* Washington, DC: World Bank.

12. Hannan, M.A., Arebey, Maher, Begum, R.A., and Hassan Basri. (2011). Radio Frequency Identification (RFID) and communication technologies for solid waste bin and truck monitoring system. *Waste Management, 31*(12), 2406-2413.

13. Karthik, M., Sreevidya, L., Nithya Devi, R., Thangaraj, M., Hemalatha, G., and R. Yamini. (2023, june 10). An efficient waste management technique with IoT based smart garbage system. *Materials Today: Proceedings*, *80*, 3140-3143.

14. Sudha, S., Vidhyalakshmi, M., Pavithra, K. (2016). An automatic classification method for environment: Friendly waste segregation using deep learning.

15. Mittal, G., Yagnik, K.B., Garg, M., Krishnan, N.C. (2016). SpotGarbage: smartphone app to detect garbage using deep learning. *the 2016 ACM International Joint*, (pp. 940-945).

16. George E. Sakr; Maria Mokbel; Ahmad Darwich; Mia Nasr Khneisser; Ali Hadi. (2016). Comparing deep learning and support vector machines for autonomous waste sorting. *ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 207–212). Beirut, Lebanon: IEEE.

17. A. Krizhevsky, I. S. (2012). *Imagenet classification with deep convolutional neural networks.* Advances in neural information processing systems.

18. P, S. (2023). Classification of Waste Materials using CNN Based on Transfer Learning. *Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation*, (pp. 29-33).

19. Yang, M. T. (2016). *Classification of trash for recyclability status.* CS229 Project Report.

20. Adedeji, O. a. (2019). Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing , 35*, 607-612.

21. Pathak, D. R. (2017). *Solid Waste Management Baseline Study of 60 New Municipalities.* Tech. Report. 10.13140/RG.2.2.11930.24006/1., Tribhuvan University, Nepal.