SETSCI Conference Proceedings
22, 34-38, 2025

International Symposium on AI-Driven Engineering Systems

June 19-20, 2025, Tokat, Türkiye

https://doi.org/10.36287/setsci.22.3.001

A Deep Learning Framework for the Identification of Distinct Stages in Diabetic Retinopathy through Retinal Image Analysis

A.D.P. Rupasinghe[1] and K.G. Samarawickrama[1*]

[1]*Department of Electrical, Electronic and Telecommunication Engineering, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka (samarawickramakg@kdu.ac.lk)*

Abstract - Diabetic retinopathy is the leading cause of blindness among diabetic patients. It occurs when the light-sensitive tissue in the retina is damaged. According to WHO statistics, the current global prevalence of diabetic retinopathy is approximately 103 million and is expected to be 161 million by 2045. The existing diagnosis methods are enriched with advanced retinal image processing and feature extraction techniques are frequently used to detect the presence of diabetic retinopathy in patients rather than focusing on early-stage detection. Therefore, detecting diabetic retinopathy at an early stage is crucial to prevent severe complications. The current study has explored the feasibility of identifying different stages of diabetic retinopathy using retinal images analyzed through deep learning algorithms. Labeled data corresponding to the five stages of diabetic retinopathy from the APTOS 2019 blindness detection dataset were preprocessed and a transfer learning approach was applied, utilizing pre-trained models from ImageNet for training. Among the transfer learning algorithms, the neural networks of ResNet_101, DenseNet_201 and EfficientNet_b0 were selected to train three DL models to select the optimal model based on test accuracy. The proposed method achieved a test accuracy of 91% using the fine-tuned EfficientNet_b0 model.

Keywords – Stages of Diabetic retinopathy, Deep learning, Training, Test accuracy, EfficientNet_b0

I. Introduction

Diabetes is a chronic medical condition in which the human body cannot produce or use enough insulin to regulate blood glucose. It is diagnosed when the fasting glucose level remains elevated above 126 mg/dL in the bloodstream [1]. There are two types of Diabetes: Type 1 can be managed with insulin injections, dietary restrictions and physical activity. However, Type 2 is chronic and exists for the rest of the person's life. Uncontrolled diabetes increases the risk of cardiovascular complications, visual impairment and blindness [1]. Diabetic retinopathy (DR) is one of the leading causes of blindness. High blood glucose levels can damage the retinal blood vasculature, leading to loss of vision. Symptoms can be subtle or difficult to detect in the early stages of the disease as the condition progresses, resulting in significant visual impairment and eventually complete blindness [2]. Based on WHO statistics, the current global prevalence of DR is approximately 103 million and is expected to be 161 million by 2045 [2].

The current research study aims to use image processing and deep learning (DL) techniques for the early detection of DR through the classification of its various stages. The APTOS 2019 blindness detection dataset comprises retinal images labelled with stages of DR (available at Kaggle.com) was employed for the study [3]. The dataset includes 3,662 retinal images for training and 1,928 retinal images for testing. Image augmentation, resizing and normalization techniques were used for the preprocessing stage to eliminate subject bias variability. The training data were input into three transfer learning DL architectures—DenseNet 201, ResNet 101, and EfficientNet_b0 which were employed from Imagenet (Image database with pre-trained DL models and parameters) to develop classification models for the distinct stages of diabetic retinopathy (DR) [4]. The three models were evaluated using the testing data, and after assessing their accuracies, the EfficientNet_b0 model demonstrated the highest performance, prompting further fine-tuning of its parameters to enhance the accuracy of classifying the distinct stages of DR.

II. Literature Review

There are two major categories of DR: non-proliferative DR (NPDR) which is an early stage of DR where the retinal blood vessels become weakened leading to swelling and haemorrhage and proliferative DR (PDR) which is a severe level of DR where abnormal new blood vessels formation on the retina due to poor circulation and oxygen supply [5]. NDPR is further divided into three stages: mild NDPR, moderate NDPR and severe NDPR. Based on the DR categorization, there are publicly available retinal image datasets such as APTOS 2019 blindness detection dataset with labelled information which can be utilized for analysing retinal features with image processing, training DL models and developing automated diagnostic tools.

Image processing algorithms are vital for the extraction of critical features and patterns, facilitating accurate diagnosis in medical imaging applications. In the context of preprocessing of retinal images techniques such as image augmentation, resizing, contrast enhancement and normalization are used for an effective feature extraction, analysis and reliable disease stage identification. Image resizing ensures uniform dimensions across images before input into machine learning (ML) and DL models, while image augmentation increases dataset diversity by generating transformed image versions. Moreover, noise reduction techniques, including Gaussian and Kirsch filters, improve image clarity, while image enhancement methods such as histogram equalization and adaptive histogram equalization enhance contrast and visual quality, highlighting essential features [6].

Feature extraction is crucial for identifying retinal areas of microaneurysms, exudates and haemorrhages in the distinct stages of DR, offering valuable insight into its progression and severity. Consequently, ML and DL models are often trained for feature extraction and pattern recognition, enabling the detection of lesions and abnormalities of the retina that may not be discernible to the naked eye. Although there are conventional ML models such as Support Vector Machines (SVM), k-nearest Neighbors (KNN) and Convolutional Neural Networks (CNN) capable of classifying the stages of DR, DL models like EfficientNet, DenseNet and ResNet have emerged as more effective alternatives for achieving higher accuracy, particularly in tasks with limited data [5]. DL models offer autonomous learning of hierarchical data representations and recognition of complex retinal patterns that indicate distinct stages of DR. These techniques are recurrently used with each network layer learning to recognize more complex features. Additionally, transfer learning has been used in some studies, where pre-trained DL models with ImageNet dataset have been used for preliminary investigations of classifying distinct stages of DR [7].

III. Methodology

A. Retinal Image Dataset

The methodology involved gathering a reliable collection of retinal images to recognize distinct stages of DR accurately. The current study has used the Kaggle website platform to access retinal image datasets, specifically the Asia Pacific Tele-Ophthalmology Society (APTOS) Blindness Detection dataset [3]. This dataset contains 3662 training images and 1928 testing images. Each retinal image has been rated by a clinician for the severity of DR on a scale from 0 to 4, where 0 represents no DR (control), 1 represents mild NPDR, 2 represents moderate NPDR, 3 represents severe NPDR and 4 represents PDR. From the training dataset, a minimum of 190 retinal images from each class were randomly extracted downsizing the dataset into 950 samples covering all the distinct stages of DR.

B. Image Preprocessing Techniques

The stages of image preprocessing techniques performed for the retinal images are explained in the following sub-sections.

1) Image Augmentation: This study explored various augmentation techniques, including rotations and flipping. The original retinal images were rotated by 90°, 180° and 270°, as shown in Fig.1, and flipping was applied along both the vertical (top-bottom) and horizontal (left-right) axes as depicted in Fig.2. Thereby, the final training retinal image dataset expanded to 5700 samples (190 x 6 x 5).

2) Image Resizing: The original retinal image size (2588x1958) of the training dataset was resized to 224x224 image size where the selected of pre-trained DL models were compatible with such resized input images. This approach eliminates the memory restrictions and potential overfitting during the model training. Larger images contain more information and may require more parameters, making 224x224 a standard size used in popular models like EfficientNet, ResNet, and Densenet [8].

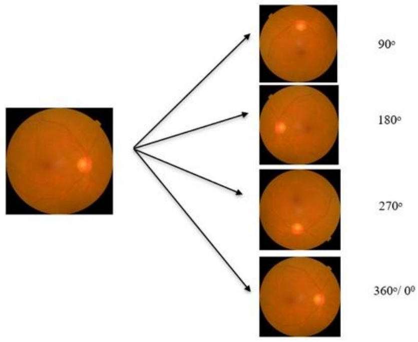

Fig. 1 Rotated Retinal Images

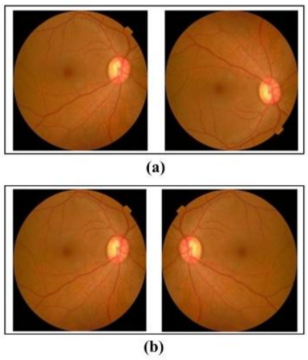

Fig. 2 Rotated Retinal Images

3) Data Normalization and Splitting: After completion of augmentation and resizing, the retinal images were normalized to match ImageNet DL models' statistical parameters to improve the performance of classification of DR stages. The mean values were (0.485, 0.456, 0.406), and the standard deviations were (0.229, 0.224, 0.225) for DenseNet 201, ResNet_101, and EfficientNet_b0, respectively [9]. Then the training dataset was divided into 80% for training and 20% for testing, ensuring models learn underlying patterns while retaining sufficient data for evaluation. Due to limited labelled data, the test dataset was used for DL model validation.

C. Model Training

Transfer learning on a pre-trained DL model on the ImageNet, enabling the identification of distinct stages of DR using the specific APTOS retinal images dataset. DenseNet_201, ResNet_101, and EfficientNet_b0 architectures were used to determine the most accurate DL model.

1) DenseNet_201: A convolutional neural network architecture that facilitates efficient feature reuse by establishing direct connections between layers. It is particularly effective in computer vision tasks, such as image classification and object recognition. By minimizing the number of parameters, promoting feature reuse, and enhancing generalization, DenseNet improves model performance and efficiency [10].

2) ResNet_101: A deep convolutional neural network comprising 101 layers, utilizing "skip connections" to address the vanishing gradient problem and facilitate the training of deeper models. It has demonstrated superior performance on datasets such as ImageNet and is integrated into deep learning frameworks like PyTorch and TensorFlow, allowing for transfer learning using pre-trained weights [10].

3) EfficientNet_b0: A variant of the EfficientNet architecture, recognized for its exceptional performance while requiring fewer parameters and computational resources. As the smallest and most computationally efficient model, with 5.3 million parameters, it has achieved state of the art results in image classification, object detection, and semantic segmentation tasks across various datasets [11].

D. Fine Tuning

Fine-tuning is an essential aspect of deep learning, wherein the weights of the pre-trained model are adjusted for a new dataset or task while keeping the weights of the earlier layers fixed. This method enhances model performance, particularly when working with small or domain-specific datasets. Depending on the task requirements, fine-tuning can be applied to all layers or a selected subset of layers.

When fine-tuning DL models such as ResNet_101, DenseNet_201, and EfficientNet_b0, hyper-parameters including weight initialization, optimizer choice, batch size, number of epochs and learning rate play a critical role in determining model performance. These settings must be carefully adjusted according to the specific application, dataset, and task to achieve optimal results.

1) Epochs: An epoch is a single pass over the entire training dataset, where the model is exposed to all the data and modifies its parameters based on the calculated loss or error. The number of epochs is a hyper-parameter which specifies how often the learning algorithm must run through the entire dataset. Each sample in the training dataset has one opportunity to change the internal model parameters during an epoch. The number of epochs is determined based on the number of stages, data complexity and availability of computational power [12]. In this project, DenseNet_201, ResNet_101, and EfficientNet_b0 classification layers were trained for 20 epochs for comparison purposes at the initial stage.

2) Learning rate: The learning rate is a crucial parameter that regulates the speed at which an algorithm adjusts its parameter values. It affects the neural network's weights and the pace at which the model's parameters are updated. A lower learning rate may slow the convergence but lead to more accurate solutions. The learning rate can be modified based on the task, dataset, and model architecture during training. When fine-tuning pre-trained models, the choice of learning rates is influenced, with smaller rates for pre-trained parameters and larger rates for output layer parameters.

In the preliminary stage, a learning rate of 0.00005 was employed for all the pre-trained DL models (DenseNet, ResNet and EfficientNet). However, it is recommended to apply a smaller learning rate to the pre-trained parameters and consequently, the learning rate for these parameters was reduced to 0.000025 during the fine-tuning process [13].

3) Weight Decay: Weight decay is known as L2 regularization (prevent overfitting), a technique that improves the generalization ability of DL models, particularly neural networks. It penalizes the size of model weights, encouraging simpler functions less prone to overfitting training data. Weight decay can be applied to any differentiable loss function and is often combined with stochastic gradient descent [14]. In this study, the classification layers of DenseNet_201, ResNet_101, and EfficientNet_b0 were trained using a weight decay coefficient of 0.1. However, it was increased to 0.15 during the fine-tuning of the EfficientNet_b0 model.

4) Batch size: The batch size in DL refers to the number of samples a neural network processes in a single pass. This parameter can influence model accuracy, convergence and model optimization. Commonly recommended batch sizes in DL such as 32, 64, and 128 samples, strike a balance between processing speed and model accuracy [15]. This study reduced the initial batch size of 128 samples to 32 during the finetuning process.

5) Training Accuracy: Training accuracy refers to the percentage of correct predictions performed by the model during the training process. Based on the hyperparameter values, selected DL models were trained, and the training accuracy percentages were achieved as follows, DenseNet_201: 59.91%, ResNet_101: 58.07% and EfficientNet_b0: 61.05% respectively. However, it was highlighted that these training accuracies are below the nominal value (typically expect training accuracies greater than 85%), thereby it was essential to follow a fine-tuning procedure to enhance the training accuracy. Due to the extensive training time, only the EfficientNet_b0 with the highest among the 3 DL models was carried out for the fine-tuning and validation.

E. Model Validation

EfficientNet_b0 was chosen for fine-tuning due to its high accuracy in the training procedure apart from DenseNet_201 and ResNet_101. As explained in the section III-D, the EfficientNet_b0 model was fine-tuned according to the adjusted hyperparameters and retrained on 950 retinal images used in the training phase. After recurring 60 epochs, EfficientNet_b0 was able to reach 95% test accuracy.

Furthermore, the model was validated from randomly selected retinal images from test data (not used for model training) from APTOS 2019 blindness detection dataset and was able to obtain 90% of accuracy from EfficientNet_b0.

IV. Results And Discussion

The transfer learning process was executed by initializing the hyperparameters for all three DL models including a batch size of 256, 20 epochs, a learning rate of 0.0005 and a weight decay of 0.1. The learning curves of accuracy and loss for DenseNet_201, ResNet_101 and EfficientNet_b0 are shown in Fig.3, Fig.4 and Fig.5 respectively. All three DL models have shown good learning curve trends. However, it was observed that the training and validation accuracies were lower (55% to 60% range) than anticipated, indicating that the model may not have learned the underlying patterns in the training data effectively. Therefore, optimizing the hyperparameters was essential during the fine-tuning procedure. Due to the extended training time required, only EfficientNet_b0 was selected for further analysis based on the highest test accuracy among the three DL models and promising learning curve trends compared to DenseNet_201 and ResNet_101.

Confusion matrix is used as a performance evaluation tool used in classification tasks to assess the accuracy of DL models. The confusion matrix of the EfficientNet_b0 model before fine-tuning is shown in Fig.6 which reflects the numerical values of true and false predictions generated for each of the distinct stages of DR. The test accuracies obtained before finetuning for each category classification are listed in Table 1. It was noted that, before fine-tuning the DL model, it was more effective in predicting the absence of DR (No DR) and the presence of DR, rather than distinguishing between the intermediate stages of DR. However, the current study aimed to distinguish the distinct stages of DR, hence the fine-tuning process was performed to improve the test accuracies of the classification.

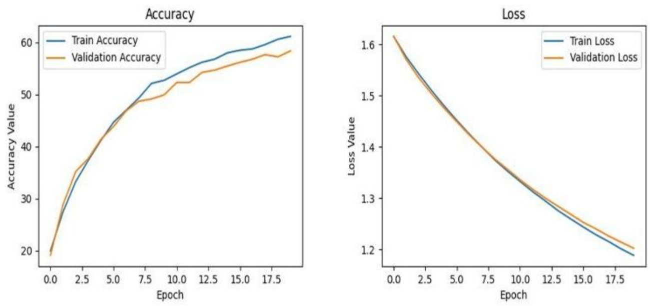

Fig. 3 DenseNet_201 Learning Curve

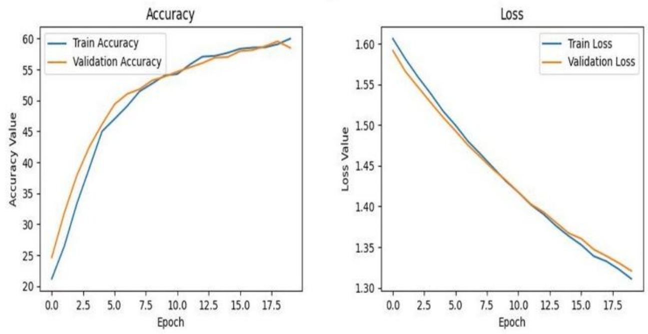

Fig. 4 ResNet_101 Learning Curve

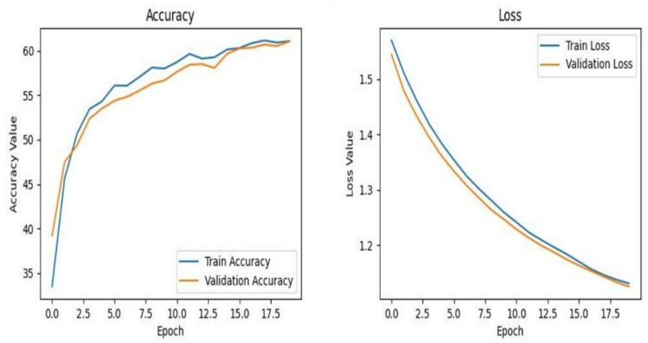

Fig. 5 EfficientNet_b0 Learning Curve

To achieve high accuracy from the EfficientNet_b0 model, hyperparameters were adjusted to increase the number of epochs from 20 to 60, reducing the batch size from 128 to 32 and the setting weight decay from 0.10 to 0.15. With this fine-tuning, the training accuracy of 95% was achieved from the EfficientNet_b0 model learning curve as depicted in Fig.7. The confusion matrix of the EfficientNet_b0 model after fine-tuning is shown in Fig.8 and the test accuracies for each distinct category of DR are tabulated in Table 2. After the fine-tuning process, all the test accuracies for distinct DR stages have reached close to 90% indicating that the trained EfficientNet_b0 model is highly effective in DR classification.

Table 1. Test accuracy of the classes before the fine-tuned model

Category	Accuracy Before Fine-tuning
No DR - 0	0.77153
Mild NPDR - 1	0.65140
Moderate NPDR - 2	0.59375
Severe NPDR - 3	0.52396
PDR- 4	0.48485

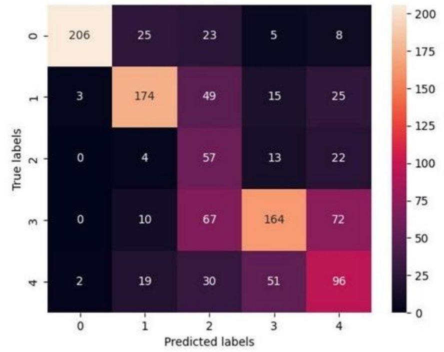

Fig. 6 Confusion matrix of EfficientNet_b0 before Fine-tuning

The validation accuracy also reached 91.4% and it was confirmed from the randomly selected testing dataset retinal images. The results of twenty randomly selected retinal images from the test dataset, actual Vs prediction outputs are tabulated in Table 3. The results demonstrate a prediction accuracy of 90%, including the ability to distinguish between the distinct stages of DR. The summary of the overall process flow in the EfficientNet_b0 DL model training and validation is depicted in Fig.9. The current study's aim was achieved through the training of the EfficientNet_b0 model which could predict the distinct stages of DR with high accuracy.

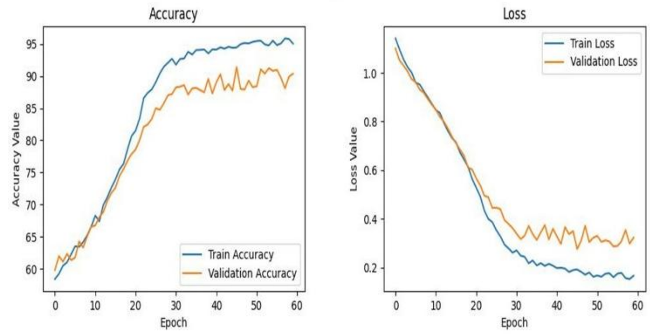

Fig. 7 Learning Curve of EfficientNet_b0 model after Fine-tuning

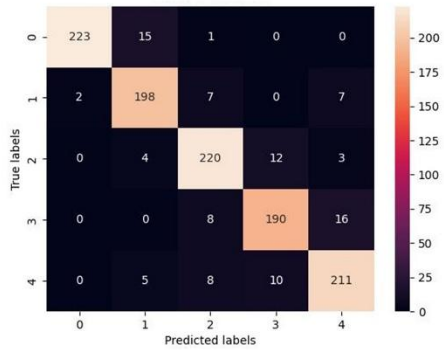

Fig. 8 Confusion matrix of EfficientNet_b0 after Fine-tuning

Table 2. Test accuracy of the classes after the fine-tuned model

Category	Accuracy After Fine-tuning
No DR - 0	0.93305
Mild NPDR - 1	0.92523
Moderate NPDR - 2	0.92050
Severe NPDR - 3	0.88785
PDR- 4	0.90171

Table 3. Actual vs predicted outputs (Accuracy of 90%)

Image Label	Actual Category	Predicted Category
0_1698571252866	No DR	No DR
0_1698571252869	No DR	No DR
0_1698571253017	No DR	No DR
0_1698571253022	No DR	No DR
1_1698571252778	Mild NPDR	Mild NPDR
1_1698571252783	Mild NPDR	Mild NPDR
1_1698571252785	Mild NPDR	Mild NPDR
1_1698571254583	Mild NPDR	Mild NPDR
2_1698571252239	Moderate NPDR	Moderate NPDR
2_1698571254118	Moderate NPDR	PDR
2_1698571254121	Moderate NPDR	Moderate NPDR
2_1698571254123	Moderate NPDR	Moderate NPDR
3_1698571255194	Severe NPDR	Severe NPDR
3_1698571255199	Severe NPDR	Severe NPDR
3_1698571262441	Severe NPDR	Severe NPDR
3_1698571267473	Severe NPDR	Severe NPDR
4_1698571252545	PDR	PDR
4_1698571252548	PDR	PDR
4_1698571260113	PDR	PDR
4_1698571260116	PDR	Severe NPDR

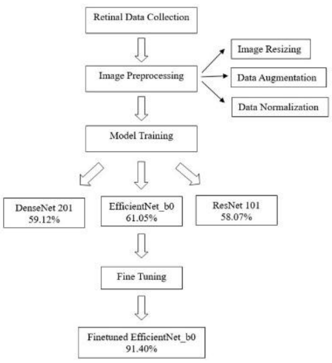

Fig. 9 Summary of the EfficientNet_b0 model Training and Validation

V. Conclusion

DR is a significant consequence in Diabetic patients which leads to visual impairments and total blindness. However, most diabetes patients are unaware of the importance of regular vision monitoring and the possibility of early-stage identification. Traditional methods of diagnosing DR, such as retinal image observation, may fail to detect the early stages of subtle symptoms without the expertise of an experienced ophthalmologist. Therefore, there is a growing need to train DL models to extract features and recognize patterns from retinal images that may not be perceptible to the naked eye and improve to early detection of DR. This study aimed to address this by implementing a DL model that accurately identifies distinct stages DR. The EfficientNet_b0 model demonstrated exceptional performance, achieving 91% test accuracy after thorough training and fine-tuning the hyperparameters. The findings of the study have shown the feasibility of early detection of DR before reaching severe stages from the EfficientNet_b0 model assessment on retinal images. Therefore, with further research work and DL model training and testing, it holds the potential to be a reliable diagnostic tool for ophthalmologists to determine the severity levels of DR. This advancement could enable diabetic patients to undergo early preventive measures, improving their long-term health outcomes. Additionally, these DL models can be integrated with retinal image acquisition devices and analysis software for faster and efficient diagnostic and treatment procedures.

Acknowledgment

The authors thank the Academic Staff of the Biomedical Engineering Degree Program, Department of Electrical, Electronic and Telecommunication Engineering, Faculty of Engineering of General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka providing valuable support and computational facilities.

References

[1] S. A. Antar *et al.*, "Diabetes mellitus: Classification, mediators, and complications; A gate to identify potential targets for the development of new effective treatments," *Biomed. Pharmacother.*, vol. 168, p. 115734, Dec. 2023, doi: 10.1016/j.biopha.2023.115734.

[2] T. Y. Wong and T.-E. Tan, "The diabetic retinopathy 'pandemic' and evolving global strategies: The 2023 Friedenwald Lecture," *Invest. Ophthalmol. Vis. Sci.*, vol. 64, no. 15, p. 47, Dec. 2023, doi: 10.1167/iovs.64.15.47.

[3] Karthik, Maggie, and S. Dane, "APTOS 2019 Blindness Detection," *Kaggle*,2019.[Online].Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection

[4] J. D. Bodapati *et al.*, "Blended multi-modal deep ConvNet features for diabetic retinopathy severity prediction," *Electronics*, vol. 9, no. 6, p. 914, May 2020, doi: 10.3390/electronics9060914.

[5] L. Qiao, Y. Zhu, and H. Zhou, "Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms," *IEEE Access*, vol. 8, pp. 104292–104302, 2020, doi: 10.1109/ACCESS.2020.2993937.

[6] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, J. Ma, and K. Wang, "Image preprocessing in classification and identification of diabetic eye diseases," *Data Sci. Eng.*, vol. 6, no. 4, pp. 455–471, Dec. 2021, doi: 10.1007/s41019-021-00167-z.

[7] N. Tsiknakis *et al.*, "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review," *Comput. Biol. Med.*, vol. 135, p. 104599, Aug. 2021, doi: 10.1016/j.compbiomed.2021.104599.

[8] D. Mandl *et al.*, "Learning lightprobes for mixed reality illumination," in *Proc. IEEE Int. Symp. Mixed Augment. Reality (ISMAR)*, Oct. 2017, pp. 82–89, doi: 10.1109/ISMAR.2017.25.

[9] "Models and pre-trained weights — Torchvision 0.17 documentation," *PyTorch*,[Online].Available:https://pytorch.org/vision/stable/models.html#classification

[10] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[11] L. Wang, "EfficientNet — An elegant, powerful CNN," *Medium*, Oct. 1, 2022. [Online]. Available: https://pub.towardsai.net/efficientnet-an-elegant-powerful-cnn-6e2a8d528ae3

[12] "What is epoch in machine learning?— UNext," *UNext*, Nov. 24, 2022. [Online]. Available: https://u-next.com/blogs/machine-learning/epoch-in-machine-learning/

[13] "What is learning rate in machine learning? The full guide," *Deepchecks*, May 27, 2024. [Online]. Available: https://www.deepchecks.com/glossary/learning-rate-in-machine-learning/

[14] A. Kumar, "Weight decay in machine learning: Concepts – Data Analytics," *Data Analytics*, Jun. 7, 2022. [Online]. Available: https://vitalflux.com/weight-decay-in-machine-learning-concepts/

[15] D. Mandy, "Batch size in a neural network explained," *deeplizard.com*. [Online].Available: https://deeplizard.com/learn/video/U4WB9p6ODjM