International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Computer-Aided Diagnostic System for Detection and Classification of Different Grades of Diabetic Retinopathy using Ensemble Learning and Deep Learning Techniques

Sharda Dhavale[1], Archana Bhamare[2], Pallavi Adke[3], Yash Bellary*[4], Pushpa Bangare[5]

Submitted:11/03/2024 **Revised**: 26/04/2024 **Accepted**: 03/05/2024

Abstract: Diabetes, being a chronic condition, possesses the capacity to instigate a global healthcare catastrophe. Diabetes is a significant etiological agent in the development of numerous diseases and health conditions. Diabetic Retinopathy (DR) is an ocular illness that damages retinal blood vessels. If not recognized early, DR weakens vision and may result in blindness. This disease can be treated and potentially cured if diagnosed and treated promptly. Manual diagnosis of this condition (by clinicians) is time-consuming and prone to inaccuracy. Integrating machine learning technology with medical science enables precise prognosis of an individual's susceptibility to DR. The proposed work describes a Computer-Aided Diagnostic (CAD) Application for DR Prediction and Classification that employs Ensemble learning techniques and a combination of two pre-trained Deep Neural Network (DNN) models (MobileNetV2, EfficientNetB0). We create an Ensemble Stacking-based model (Ensemble Model) and an Ensemble Bagging-based model (Bagged Model), in addition to a regular Convolutional Neural Network (CNN). We study and analyse the performance of the CNN model, and DNN models (MobileNetV2, EfficientNetB0) individually, along with the Ensemble Model and Bagged Model. The Bagged Model outperforms the other models, with a training accuracy of 87.11%. Hence, this work determines the Bagged Model's potential as a viable instrument for detecting DR and is used in the development of an efficient and effective Application for Predicting Diabetic Retinopathy.

Keywords: *Bagging, Ensemble Learning, EfficientNetB0, Deep Learning, Diabetic Retinopathy, MobileNetV2.*

1. Introduction

The rising global prevalence of diabetes, expected to reach 592 million by 2035, represents a major health challenge [1]. Diabetic retinopathy (DR) is a retinal ailment inducted by diabetes. It is the leading cause of blindness worldwide, especially in developed countries [2]. According to the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), approximately 28.5 percent of Americans with diabetes aged 40 years or older have retinopathy. This is 3.7 million cases, which is double the number of people with retinopathy in 2000. We can see that between 12,000 and 24,000 new cases of blindness due to diabetes are reported each year - these numbers have caused government institutes such as the Centers for Disease Control and Prevention to label this disease a growing epidemic [3].

[1]*Research Scholar, Sinhgad college of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.*
ORCID ID: 0000-0002-7951-9395
[2]*Pimpri Chinchwad College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.*
ORCID ID: 0000-0001-8630-9671
[3]*Pimpri Chinchwad College of Engineering and Research, Savitribai Phule Pune University, Pune, Maharashtra, India.*
ORCID ID: 0000-0001-8630-1549
[*4]*Pimpri Chinchwad College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.*
ORCID ID: 0009-0008-5836-6706
[5]*SKN College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India.*
ORCID ID: 0000-0001-8326-6346
**Corresponding Author Email: bellaryyash23@gmail.com*

The DR has two major types: the Non-Proliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR) [4]. The early stages of DR are known as NPDR, and they are further categorized into Mild, Moderate, and Severe. The mild stage includes one micro-aneurysm (MA), which is a little circular red dot at the end of blood vessels. In the Moderate stage, the MAs burst into deeper layers, resulting in a flame-shaped hemorrhage in the retina.

The severe stage includes more than 20 intraretinal hemorrhages in each of the four quadrants, as well as confirmed venous bleeding and extensive intraretinal microvascular anomalies. PDR is the advanced stage of DR that results in neovascularization, which is the natural development of new blood vessels in the form of functioning microvascular networks on the retina's inner surface [5]. Figure 1 visually presents the different stages of DR.

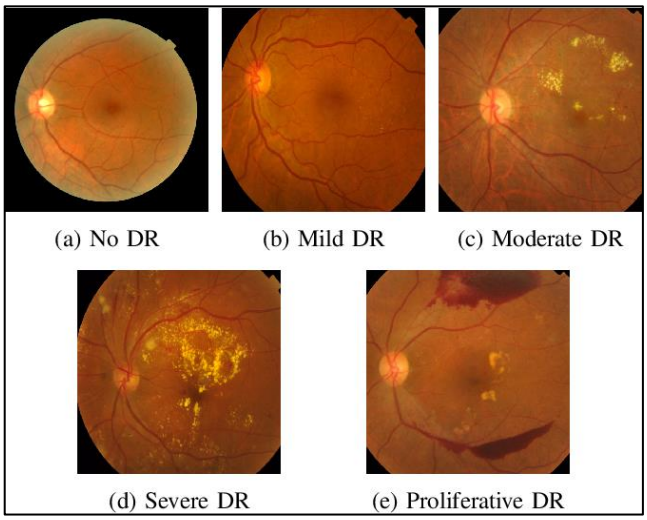

Fig. 1. Fundus Retinal Images of different stages of Diabetic Retinopathy (DR)

The stages of DR from Non proliferative DR to proliferative DR can be descriptively classified using simple international DR classification criteria, as shown in Table 1.

Table 1. Descriptive classification of different stages of DR

Sr. No.	Disease Severity Level	Abbreviations	Observational Findings
1.	No Apparent DR	NADR	No abnormalities
	Mild Non-Proliferative DR	Mild NPDR	Microaneurysms only
3.	Moderate Non-Proliferative DR	Moderate NPDR	More than just microaneurysms but less than severe Non proliferative DR
4.	Severe Non-Proliferative DR	Severe DR	Any one-off following • More than 20 intraretinal haemorrhages in each of 4 quadrants. • Definite venous beading in 2+ quadrants. • Prominent intraretinal microvascular abnormalities in 1+ quadrants. • No signs of proliferative DR
5.	Proliferative DR	PDR	One or more of following: • Neovascularization • Vitreous/ Preretinal haemorrhages.

The current gold standard for detecting diabetic retinopathy, ophthalmoscopy, is insufficient for screening due to its insensitivity to the early changes in the eye caused by diabetes [6]. According to the NIDDK, ophthalmoscopy is insensitive to the early changes in the eye caused by diabetes and is therefore insufficient for screening [[7]. With the rising prevalence of diabetes worldwide, it is crucial to detect retinopathy early for effective treatment and prevention.

Computer-assisted diagnosis has gained popularity in many parts of medicine, especially in the analysis of chronic diseases. The large diagnostic power of artificial-based systems makes them an ideal choice for automation in detecting DR [8]. The ultimate goal of our research is to facilitate the development of automated tools that significantly improve the speed and accuracy of DR diagnosis. This will reduce the immense number of people affected by this disease, making the diagnostic method more accessible to patients. The proposed CAD application for DR prediction and classification is a step towards achieving this goal. The application is designed to be user-friendly and efficient, making it a valuable tool in the fight against this growing epidemic.

The subsequent content constitutes the paper's outline. Recent work on this subject and the literature review are highlighted in Section 2. In addition to describing the methodology utilized in creating and developing the proposed model, Section 3 provides theoretical details regarding the algorithms and processes. In Section 4, the outcomes and performance of the proposed model are assessed. In addition to presenting a concluding statement and final evaluation of the research findings, Section V provides the concluding statement of the study and investigates possible future applications of the designed model.

2. Literature Review

N. Sikder et. al. [9] described a novel ensemble learning-based method to determine DR's severity on the APTOS 2019 BD dataset wherein first they employed some image pre-processing techniques, next they identified the most important features in the image using Genetic Algorithm (GA) finally hyper-tuned the ensemble-based algorithm i.e. the XGBoost algorithm to provide the best performance on the created feature set. Their method provided a classification accuracy of 94.20%. I. Odeh, M. Alkasassbeh, et al.,[10] developed an ensemble classification model using popular classification algorithms such as Random Forest (RF), Neural Network (NN), and Support Vector Machine

(SVM) as base classifiers and merged their output using meta-classifier to produce a final prediction for classification of different stages of DR. Their proposed model achieved an accuracy of 75.1% in performing this operation. Similar to this, G.T. Reddy, S. Bhattacharya, et al.,[11] developed an ensemble model using an RF classifier, Decision Tree (DT) Classifier, Adaboost Classifier, K-Nearest Neighbor (KNN) classifier and Logistic Regression (LR) classifier as the baseline classification algorithms, their model produced an accuracy score of around 80%. E. AbdelMaksoud et. al., [12] developed a comprehensive system for the early detection of DR with emphasis on extracting seven of the important pathological features used for the identification of DR using segmentation techniques and a multi-label SVM. This system generated an average accuracy of 89.2%, sensitivity of 85.1%, specificity of 85.2%, positive predictive value of 92.8%. S. Gayathri, et al., [13] designed a lightweight CNN for feature extraction and employed some of the popular classification algorithms like SVM, AdaBoost, Naive Bayes, RF, and J48 for model evaluation and found the J48 outperformed all the other classifiers with an accuracy of 99.89% for binary classification and 99.59% for multi-class classification.

In their study M. Akil et. al., [14] researched the various deep CNN models for the detection of pathological signs of DR in the captured fundus images and found that among various deep CNNs InceptionV3 architecture obtained an impressive accuracy of 98.6%. They also found that with an emphasis in recent times on using smartphones for the detection of retinal abnormalities, the Xception model generated an overall accuracy of 94.6% in DR detection. L. P. Cen et. al., [15] developed a comprehensive deep neural network for the automated detection of 39 of the most common retinal abnormalities from fundus-captured images. With the proposed architecture they were able to achieve a frequency-weighted average F1 score of 0.923, sensitivity of 0.978, specificity of 0.996, and area under the receiver operating characteristic curve (AUC) of 0.9984 for multi-label classification and they were able to make the model reach the level of an average retinal specialist. Kannan, Rajeswari et. al.,[16] in the paper developed a custom CNN model based on the study that they had carried out. This model showed a satisfactory accuracy of 70.30% which was better than the ResNet network but was outperformed by the DenseNet network. They also mentioned that the data preprocessing step improved the accuracy of the networks by up to 5%. This concludes that the data preprocessing steps play an important role in the model's accuracy and also, the deep CNNs outperform other ML techniques. Sahlsten, J., et al., [17] in their research utilized a deep learning system based on the Inception-v3 architecture, pre-trained on the ImageNet dataset, for diabetic retinopathy and macular edema grading. The AUC of the best model was reported to be 0.987, comparable to another recent study. The specificity of the model was reported to be 0.909 at 0.960 sensitivity, although another study reported a higher specificity of 0.923 at 0.971 sensitivity. Oh, K., Kang, H.M., Leem, D., et al., [18] propose a diabetic retinopathy detection system based on ultra-wide-field fundus photography and deep learning. The system requires automatic segmentation of the ETDRS 7SF (Early Treatment Diabetic Retinopathy Study 7-standard field) to remove undesirable components such as eyelashes and skin. The DR detection system based on ETDRS 7SF images performs significantly better than that based on ETDRS F1-F2 images in terms of specificity. Wejdan L. Alyoubi et. al., [19] in the methodology involved exploring deep learning techniques, discussing fundus retina datasets, and presenting performance measures. Automated screening systems employing deep learning, particularly convolutional neural networks (CNN), expedite diabetic retinopathy (DR) detection and classification, enhancing early diagnosis and treatment. Hamid Safi et. al., [20] aimed to offer a comprehensive overview of the existing landscape in subclinical DR detection research. The significance of early diabetic retinopathy (DR) detection using diverse biomarkers and techniques emphasizes the need for timely intervention to prevent irreversible visual impairment. Ongoing efforts focus on developing reliable methods for early DR recognition. S. Qummar et al., [21] employed an ensemble approach involving five deep Convolutional Neural Network (CNN) models (Resnet50, Inceptionv3, Xception, Dense121, Dense169) trained on the Kaggle dataset of retina images. Through fine-tuning and stacking, the ensemble model successfully encoded rich features, achieving improved classification for different stages of Diabetic Retinopathy. Bilal A et. al., [22] introduced a novel approach for Vision-Threatening Diabetic Retinopathy (VTDR) detection using the Hierarchical Block Attention (HBA) and HBA-U-Net architecture. Employed a hybrid CNN-SVD model for feature extraction, followed by classification with an Improved Support Vector Machine (ISVM), achieving outstanding accuracy of 99.18%, sensitivity of 98.15%, and 100% specificity on the IDRiD dataset.

Michael D. Abràmoff MD et. al., [23] analyzed two algorithms, EyeCheck and Challenge2009, for their performance in detecting diabetic retinopathy (DR) lesions. While both achieved similar AUC values, combining their outputs led to increased sensitivity without compromising specificity. Despite reaching an AUC of 0.86, further improvements beyond this limit were deemed unattainable due to the dataset's characteristics and human expert readings. Borys Tymchenko1 et. al., [24] discussed the effectiveness of convolutional neural networks (CNNs) in

classifying DR stages and highlighted utilizing transfer learning techniques with pre-trained models like InceptionNet V3, ResNet50, XceptionNet, DenseNet, and VGG. Challenges such as dataset inconsistency and model architecture refinement were acknowledged, and they emphasized the need for robust deep-learning algorithms in DR diagnosis. Mishra, S. et.al, [25] in the paper highlighted various approaches to diabetic retinopathy (DR) detection using deep learning. Techniques such as transfer learning, ensemble methods, and novel CNN architectures like DenseNet and VGG16 were explored.

Research on DR categorization can be divided into binary classification and multi-class categorization models. However, most studies fail to categorize all stages of DR, particularly early ones. Traditional machine learning classification algorithms and deep learning techniques are used for classification and prediction. While classification algorithms can produce satisfactory results for binary classification, they lack proper categorization of different stages due to their limited learning ability. Deep learning models are effective in extracting strong feature maps from data, but their high computational power requirement and long training duration limit their efficiency. There is currently no comprehensive computer-aided system for predicting DR using retinal image scans.

3. Methodology

The design stages and development process undertaken for DR prediction and classification are elaborated upon in this section. A description of the design and block diagram of the proposed system is provided. Additionally, the dataset's specifics are disclosed. A detailed discussion follows regarding the performance analysis evaluation parameters.

3.1. Dataset Description

The Asia Pacific Tele-Ophthalmology Society 2019 Blindness Detection (APTOS 2019 BD) dataset's fundus images are used in the study [9] . 3662 samples were gathered from numerous people in rural India for this image dataset. The dataset was arranged by India's Aravind Eye Hospital. Using fundus photography, the hospital experts journeyed to isolated regions of India to obtain samples of retinal images. These fundus photos were taken over an extended period of time in a variety of settings and scenarios. A team of medical professionals with training later examined and identified the collected samples using the International Clinical Diabetic Retinopathy Disease Severity Scale (ICDRSS). The APTOS 2019 BD samples are evaluated as 0-no DR, 1-mild DR, 2-moderate DR, 3-severe DR, and 4-proliferative DR according to the scale scheme, as indicated in Table 2. The first group includes the healthy retinal samples free of diabetic retinopathy. Slightly greater retinal damage is represented by each of the latter groups than by the former. The final class of samples, known as proliferative DR, includes those with vitreous or pre-retinal haemorrhage.

Table 2. Dataset attributes

Grade/ Level of DR	Stages of DR	Number of samples of the category
0	No DR	1805
1	Mild DR	370
2	Moderate DR	999
3	Severe DR	193
4	Proliferative DR	295

The spread of each category in the dataset is visualized in Figure 2.

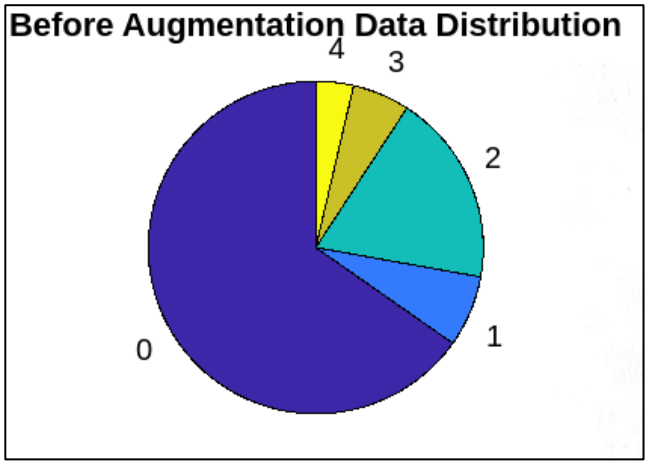

Fig. 2. Pie chart showing the spread of each category of data in the dataset

From Figure 3, it can be seen that an imbalance exists between the categories in the dataset. Where category 0 has the highest number of samples and category 4 has the least number of samples. This imbalance can affect the system's training performance and to enhance the accuracy of the ML system, this imbalanced dataset must be transformed into a balanced one [21].

3.2. Flowchart

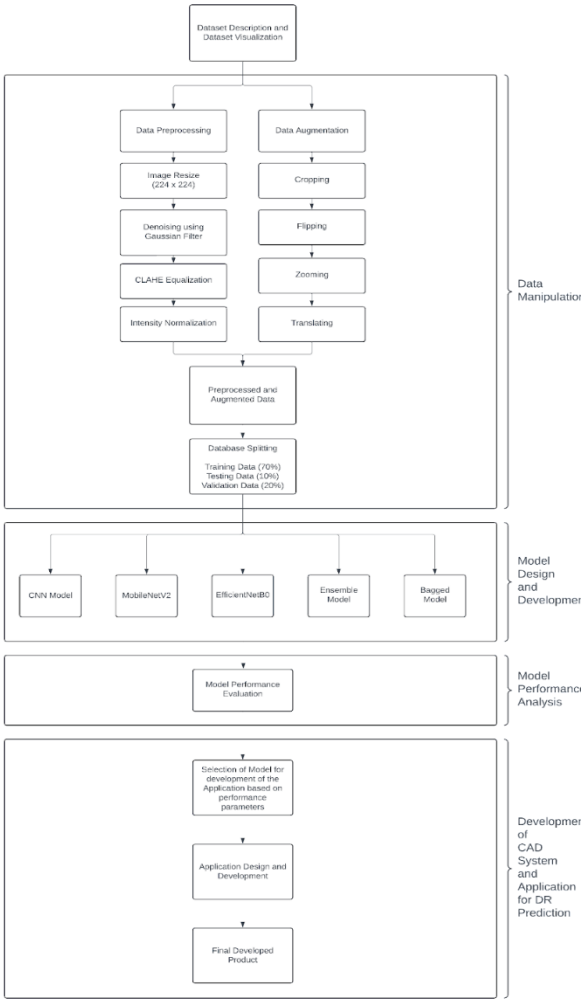

Fig. 3. Flowchart

The methodology implemented in the development of the CAD application for DR prediction is illustrated in the flowchart, as shown in Figure 3. The initial step of the process involved understanding the dataset and gathering information about the characteristics of the dataset and the spread of data in it. The next step involved the manipulation of data to alter it according to our model's requirements. This involved the steps of dataset processing wherein the size of the images was updated, and various transformations and normalization techniques were applied to enhance the image attributes, this allowed the model to extract better feature maps from the images. Since there exists an imbalance in the dataset between the category of images as shown in Figure 2, the steps of augmentation of the dataset are performed. With the use of augmentation techniques like cropping, flipping, zooming, and transforming new image samples are generated of categories that lack sufficient data samples. Once the data was ready for training the ML models, the next step involved designing the architecture of the models. After extensive research and study of recent work, the Deep learning models and the Ensemble techniques were finalized for the task of DR prediction and classification. Models like CNN, Ensemble Model, and Bagged Model were designed and developed. The trained DNNs like MobileNetV2 and EfficientB0 were modified to fit our requirements. Then the performance of each model was analyzed by generating various performance metric scores which helped in the evaluation of the models. The best-performing model was selected for the development of the final application. The layout and design of the User Interface of the application were created and the functionalities of the system were defined. Once all these steps were completed the final application was ready for prediction and grading of DR.

3.3. Data Manipulation

3.3.1. Data Pre-processing

The shape, size, and quality of the image samples in the collection were not all the same. To improve the quality of the dataset, the photos underwent data processing operations. MATLAB software was utilized for the data processing stage. Image examples with uniform 224x224 dimensions were created using MATLAB image editing methods such as "resize ()". They were divided into RGB layers once the image was resized, enabling more effective manipulation of the image's characteristics. We used a Gaussian filter to remove noise from the retinal images in our dataset because medical data is often noisy. Utilizing the MATLAB function "imgaussfilt()," a 2D Gaussian smoothing filter was applied to the picture. This improves the picture samples' numerous pathological characteristics. Using the Contrast Limited Adaptive Histogram Equalization (CLAHE) function, the image's contrasts were enhanced. The analogous MATLAB function was 'adapthisteq()'. Lastly, the 'imadjust()' function was used to transfer the intensity values in the image across the specified range of intensity values, standardizing the intensity range across the image to improve the image quality. The enhancement in the image quality after the data preprocessing on categories is evident in Figure 4 and 5.

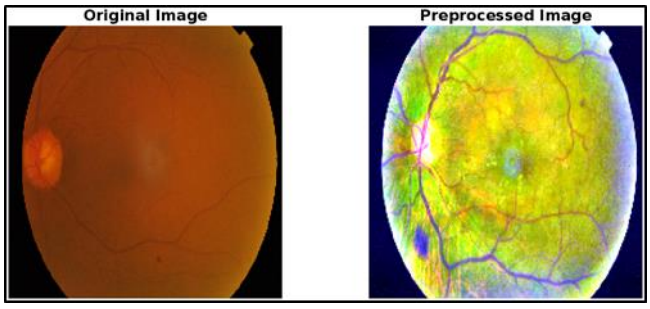

Fig. 4. Enhancement of Image sample after preprocessing operation (Category 1)

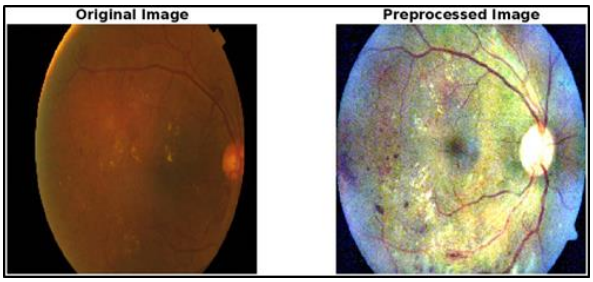

Fig. 5. Enhancement of Image sample after preprocessing operation (Category 2)

3.3.2. Data Augmentation

It is illustrated from the study conducted by S. Qummar. al., [21]that an imbalance in the dataset affects the performance of the models, and shown in Figure 3, our dataset is dominated by the samples from category 0. To address this data inconsistency, we used a variety of data augmentation approaches to increase the number of image samples from each category in the dataset. MATLAB functions such as cropping, flipping, zooming, and translating are applied to the picture sample to create new samples. These newly generated images after augmentation are illustrated in Figure 6.

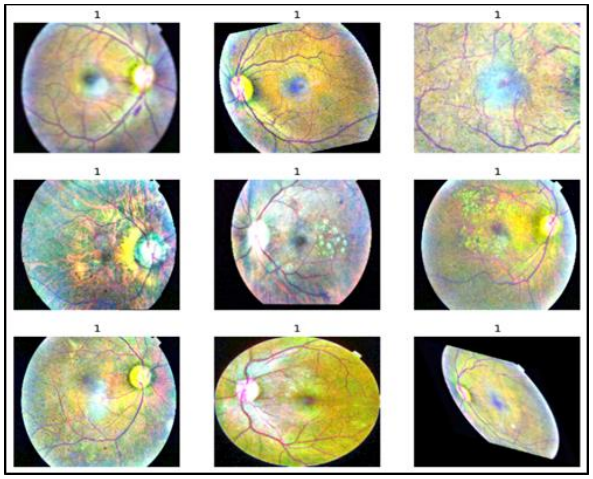

Fig. 6. Newly generated image samples from Data augmentation for Category 1

The change in the number of image samples of each category before and after augmentation is listed in below table 3.

Table 3. Variation in Dataset before and after augmentation

Grade/ Level of DR	Before Augmentation		After Augmentation	
	Number of samples	Percentage of total samples	Number of samples	Percentage of total samples
0	1805	49.2%	3492	31.58%
1	370	10.10%	2170	19.62%
2	999	27.28%	1799	16.27%
3	193	5.27%	1799	16.27%
4	295	8.15%	1799	16.27%
Total	3662	100	11059	100

Thus, after augmentation of the data, it can be seen that the imbalance between the categories in the dataset is reduced and a more evenly distributed dataset is generated. The spread in the data samples can be visualized in the below figure 7.

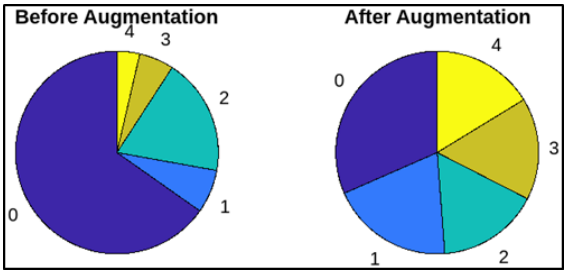

Fig. 7. Pie chart showing spread of data, before and after augmentation.

3.3.3. Data Splitting

After preprocessing and augmenting the data, the next step is to divide it into training, testing, and validation datasets. This is significant since it allows us to evaluate the model's performance on both labeled and unlabeled data. The classification and prediction performance can both be evaluated. The data is divided into three categories: training, validation, and testing, with a 70:20:10 split. This splitting procedure is carried out using the Sci-kit learn package's 'train_test_split()' function. The visualization of the data split can be seen in Figure 8.

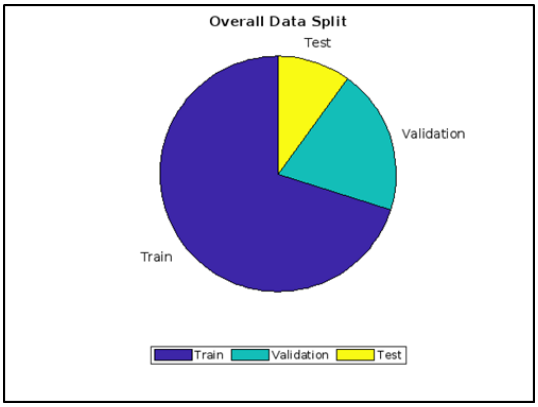

Fig. 8. Pie Chart showing Data Split

3.4. Model Design and Development

3.4.1. CNN Model

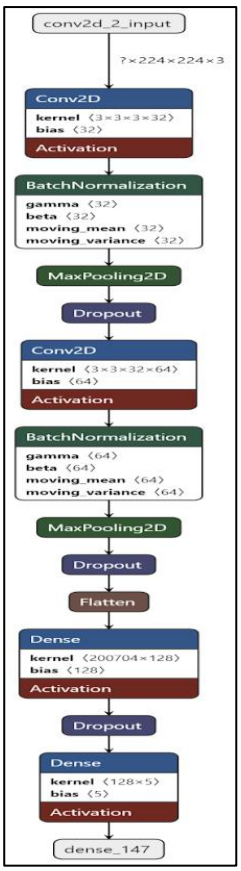

Fig. 9. Architecture of CNN Model

The design and architecture of the CNN model are shown in Figure 9. The CNN model has a depth value of 4, depicting that the model has four layers of connected neural networks. The first and second layers are convolutional layers which contain 32 neurons and 64 neurons respectively; they use the ReLU activation function and max pooling layer for extracting feature maps from the images. We apply normalization and dropout to the layers to prevent overfitting. The third layer and the fourth layer are dense layers with linear operation, which output the prediction of the grade of DR in the image sample. In this way, the CNN model's architecture is designed, and the model is implemented using TensorFlow, the open-source ML package developed by Google.

3.4.2. MobileNetV2

MobileNetV2 is a convolutional neural network architecture that is lightweight and optimized for mobile and embedded vision applications. It was trained using the ImageNet database. It uses depth wise separable convolutions to minimize computational complexity while retaining excellent accuracy. MobileNetV2 is made up of a series of layers, beginning with a conventional convolutional layer and progressing to bottleneck layers. Each bottleneck layer consists of a depth wise convolution, a pointwise convolution, and shortcut connections. This architecture enables MobileNetV2 to strike a fair balance between latency, size, and accuracy. These characteristics make it an appropriate alternative for doing DR prediction and classification on devices with minimal computational capability. The final output layer of the fully linked layers is updated in the architecture to produce a multi-class classification and prediction result. The use of the average pooling layer and normalization layer is done to prevent overfitting. Also, during the training of the model, the use of the 'ReduceLRonPlateau()' callback function from the TensorFlow package is done to modify the learning rate while training the model.

3.4.3. EfficientNetB0

EfficientNetB0, a member of the EfficientNet family of models trained on the ImageNet database, use compound scaling to balance network depth, width, and resolution. This scaling enables EfficientNet models to attain state-of-the-art accuracy with substantially fewer parameters and FLOPs (Floating Point Operations) than competing architectures. EfficientNetB0 begins with a stem convolutional layer, followed by several blocks that repeat in a same fashion. Each block is made up of depth- and point-wise convolutions that are optimized for squeeze and excitation. The number of layers, width, and resolution are tuned to maximize performance and efficiency, making it effective for anticipating DR and producing findings quickly.

Similar to the MobileNet, the architecture of the EfficientNet is modified in the final output layers to make it accustomed to our requirements.

3.4.4. Ensemble Model

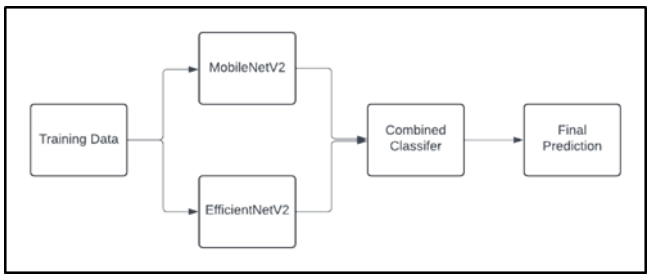

Fig. 10. Block Diagram of Ensemble Model

The block diagram of the Ensemble Model for DR prediction is shown in Figure 10. The first layer consists of a stacked classifier developed with the MobileNetV2 and EfficientNetB0 DNN models. This model generates feature maps for the first and second models using the ensemble learning stacking technique. Later, the characteristics from both models are concatenated to create new feature maps in the combined classifier layer, which include features

retrieved from both models. The retrieved features are then used to construct the DR's final predictions. This ensemble model was created using the TensorFlow package's concatenation of layers function. The output layer used average pooling and linear operations to estimate the grade of DR.

3.4.5. Bagged Model

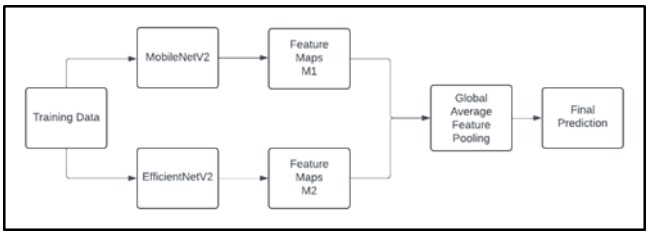

Fig. 11. Block Diagram of Bagged Model

The block diagram of the designed ensemble model using the bagging technique i.e. Bagged model is shown in Figure 11. Bagging, often referred to as bootstrap aggregation, is an ensemble learning method that enhances the prediction performance of a machine-learning model and produces a stable model by combining the advantages of aggregation and bootstrapping. The two DNN models are trained independently using the Bagged model, which produces their respective feature maps, M1 and M2. The outcomes from both models are then integrated by averaging and eliminating the weak features using the bagging technique. Ultimately, the model is trained using these robust features to produce the DR severity level grade.

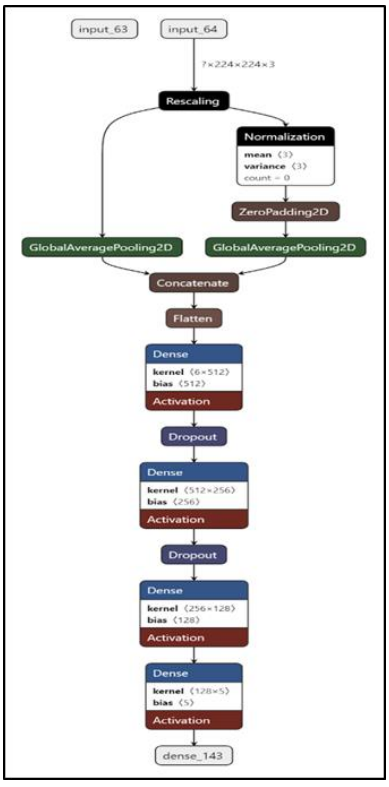

Fig. 12. Architecture of the designed Bagged Model

The architecture of the Bagged model is visualized in Figure 12. Wherein, the use of the Global Average pooling layer is done to generate aggregate feature maps canceling out the weak features. The ReLU activation function is used to generate feature maps and finally, the dense layer outputs the linear prediction of the DR class.

3.5. CAD Application for Prediction of Diabetes

3.5.1. Application Layout

Fig. 13. Application Home Screen

Fig. 14. Application Prediction Output Screen

Figure 13 and 14 illustrates the layout design of the application and the working of the application in the prediction of the different grades and levels of DR. The application is developed using the Flask server. The Bagged Model is used in the backend for predicting the category of DR from input images. The front end consists of a design layout implemented using HTML, CSS, and Bootstrap framework.

3.5.2. Application Working

When the application is opened, the user is welcomed to the home screen of the app which contains an input form where the user can upload the retinal image from which they desire to predict the level of DR. Once the image is uploaded, the user must press the ‘Predict’ button to initiate the process of prediction. After the button is pressed, the form makes a call in the backend to the Bagged Model to generate the output label. The model performs the steps of resizing and preprocessing on the input image and then generates the

final output based on the trained features. This output is then returned to the front end of the application along with the input image.

Finally, the output prediction screen is presented to the user which contains the image input by the user and the predicted label with the information about the level of DR and its short description. In this way, the CAD system for Prediction of DR is developed.

4. Results and Discussions

Comparison tables are developed for evaluating the training, testing, and validation dataset performance of each developed model. Bar plots are generated to showcase the comparison of output values of each algorithm concerning the different performance parameters.

4.1. Comparison of average accuracy and loss of each model

The comparison of average values of Training, Testing, and Validation accuracies and losses of each of the developed models is tabulated in the below table 3. From the table, we analyze that the Bagged model shows the best performance in terms of accuracy and generates the least value of loss for training, testing, and validation datasets. These findings indicate that the Bagged model can generate strong feature maps from the training data and apply these generated features to accurately predict the level of DR in testing and validation datasets while maintaining a low loss value and avoiding overfitting.

Table 4: Comparison of average accuracy and loss of developed models over train, test, and validation dataset

Models	% of Average Accuracy over epochs (values in %)			Average value of Loss over epochs		
	Train	Validation	Test	Train	Validation	Test
CNN	40.46	27.21	33.75	1.79	1.6	2.49
EfficientNetB0	53.22	38.81	43.53	1.58	1.51	1.49
MobileNetV2	52.8	37.2	37.27	1.58	1.51	1.63
Ensemble Model	73.43	67.34	51.43	0.69	0.84	1.16
Bagged Model	**87.11**	**74.17**	**68.71**	**0.44**	**0.69**	**0.89**

4.1.1. Visualization of the average % accuracy and average loss for Training data

Figure 15 shows the variation in accuracy and loss values between the models during the training process. From the plot, it is evident that the DNNs perform better when combined in a single model.

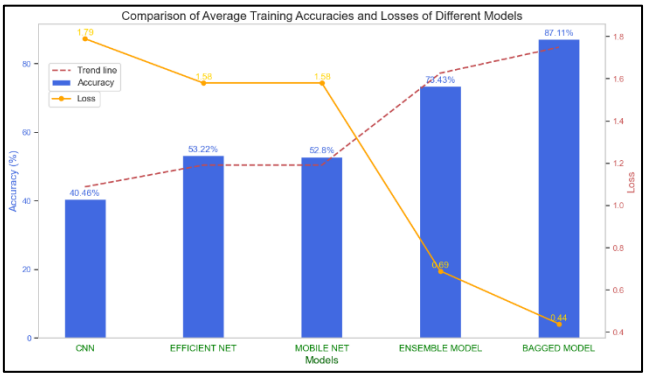

Fig. 15. Comparison plot for accuracy and loss performance on training data

4.1.2. Visualization of the average % accuracy and average loss for Validation data

In Figure 16, the accuracy and loss of models for validation data are compared. These values indicate that the models were able to extract features from training data and were able to use these features to map them to the appropriate DR class and generate predictions.

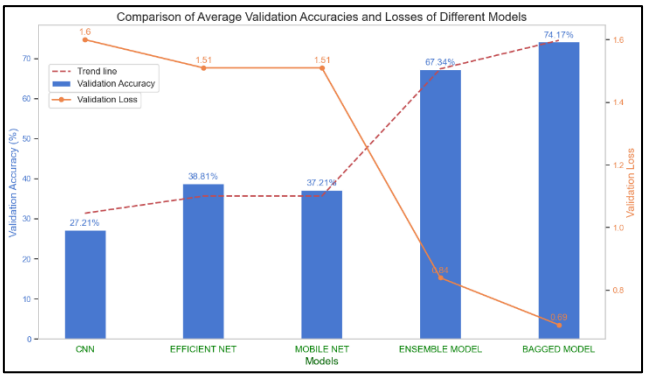

Fig. 16. Comparison plot for accuracy and loss performance on validation data

4.1.3. Visualization of the average % accuracy and average loss for Testing data

Figure 17, assesses and visualizes the performance of test data on the developed models. This helps us make sure that the models do not overfit the training data and can generate predictions with proper accuracy.

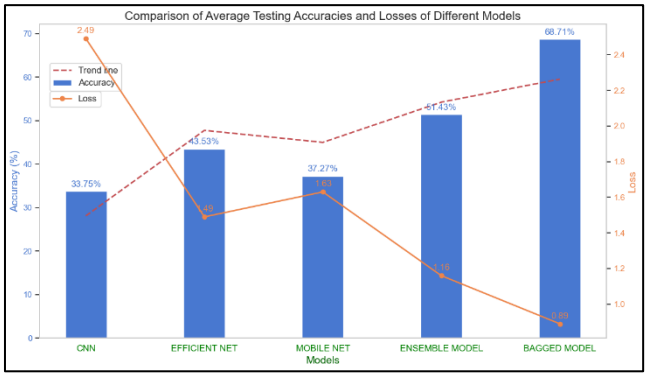

Fig. 17. Comparison plot for accuracy and loss performance on testing data

4.2. Evaluation using Performance Parameters

The parameters like precision, recall, and F1 score are used to analyze the performance of individual models on training data. Table 5 assesses the performance of each model based on the mentioned parameters. This assessment examines the performance of the models based on the predictions generated by them.

Table 5. Evaluation of models using performance parameters

Performance Parameters	CNN Model	EfficientNetB0 Model	MobileNetV2 Model	Ensemble Model	Bagged Model
Precision	0.45	0.68	0.61	0.70	0.85
Recall	0.50	0.66	0.59	0.63	0.61
F1-Score	0.47	0.65	0.60	0.76	0.69

Figure 18 presents a comprehensive comparison of the performance of all the models for each evaluation parameter.

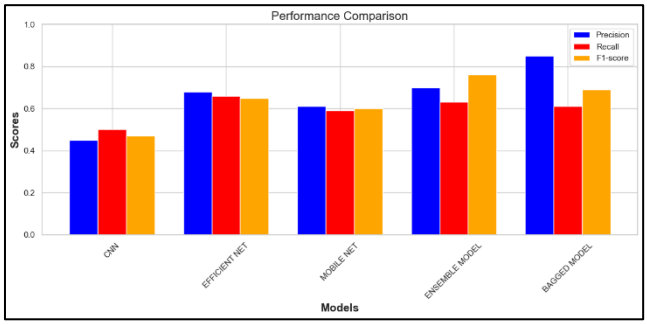

Fig. 18. Comparison of model performance using evaluation parameters

From Figure 18, it is evident that the Bagged Model and Ensemble model have superior performance in comparison to the other models. This indicates that the Bagged Model and Ensemble model correctly identify the classification of DR (Recall) and generates the appropriate prediction output (Precision). Furthermore, they incorrectly classify a smaller number of cases (F1) attesting to their robustness. These findings make them an appropriate choice for development and implementation in the CAD application for DR prediction.

4.3. Training Efficiency of Model over epochs

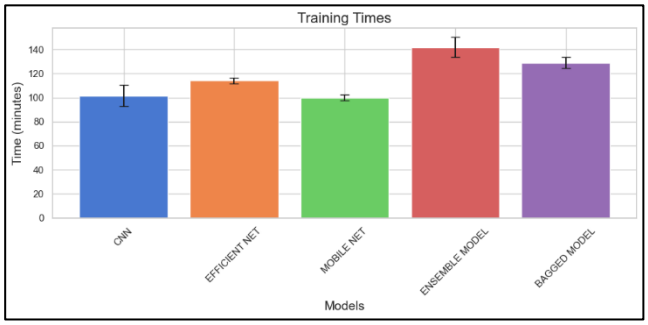

Fig. 19. Comparison plot of training times of each model

When evaluating the model's performance, training time is a significant factor. This is due to our desire to create models that are accurate and efficient in extracting features. The application will be able to show the outcomes to the user quickly and accurately if the model can train in a reasonable amount of time and produce accurate results. The training times of each model are contrasted in Figure 19. Although Table 5 makes it clear that the CNN model performs quite poorly, we can still observe that it has the shortest training time. The Bagged Model has a very high accuracy score and can train more quickly than the Ensemble model. These points make it evident that the Bagged model is efficiently and effectively able to predict DR.

5. Conclusion

This work evaluated how Deep learning algorithms can predict DR and developed a CAD Application that can predict DR in a patient with accuracy and precision. The developed Bagged model, a mix of DNNs like EfficientNetB0 and MobileNetV2 using ensemble methodology, shows promising results. For the training data, the Bagged model showed the highest accuracy of 87.11%, followed closely by the Ensemble Model at 73.43%. The overall performance of the Bagged model in the evaluation parameters is also better than the individual DNNs, which signifies that the model makes better predictions, better classifications, and substantially better coverage of the dataset than all other baseline DNNs. The EfficientNetB0 and MobileNetV2 models showed

accuracies of 53.22% and 52.8% respectively. The developed model also surpasses the performance in comparison to the previous research conducted using the same dataset. The Bagged model with an accuracy of 87.5%, exceeds the performance of models proposed by I. Odeh, and M. Alkasassbeh [10] where an accuracy of 75.1% was generated. The model also surpassed the performance of the system developed by G.T. Reddy and S. Bhattacharya [11] where an accuracy score of around 80% was generated. The developed model also significantly outperforms the system developed by Kannan and Rajeswari [16] where an overall accuracy of 70.3% was reported using ResNet DNN.

These findings imply that the prediction accuracy of individual DNNs is enhanced when combined, as shown in the Bagged model. The developed CAD Application for DR Prediction shows that machine learning algorithms can be used as practical tools for forecasting DR and can help in the timely diagnosis and prediction of DR in a patient.

References

[1] S. Jan, I. Ahmad, S. Karim, Z. Hussain, M. Rehman, and M. A. Shah, ‘‘Status of diabetic retinopathy and its presentation patterns in diabetics at ophthalmology clinics,’’ J. Postgraduate Med. Inst. (Peshawar-Pakistan), vol. 32, no. 1, pp. 24–27, Mar. 2018.

[2] E. A. Lundeen, Z. Burke-Conte, D. B. Rein, "Prevalence of diabetic retinopathy in the US in 2021," JAMA, vol. 2023. jamanetwork.com, 2023. [HTML]

[3] A. X. Chan, J. J. McDermott IV, T. C. Lee, G. Y. Ye, et al., "Associations between healthcare utilization and access and diabetic retinopathy complications using All of Us nationwide survey data," PloS one, vol. 2022, journals.plos.org, 2022. plos.org

[4] W. R. Memon, B. Lal, and A. A. Sahto, ‘‘Diabetic retinopathy,’’ The Prof. Med. J., vol. 24, no. 2, pp. 234–238, 2017.

[5] A. H. A. Ghamdi, "Clinical predictors of diabetic retinopathy progression; a systematic review," Current diabetes reviews, 2020. ingentaconnect.com

[6] C. J. Flaxel, R. A. Adelman, S. T. Bailey, A. Fawzi, J. I. Lim, et al., "Diabetic retinopathy preferred practice pattern®," Ophthalmology, 2020. [Online]. Available: aaojournal.org. aaojournal.org

[7] M. Kropp, O. Golubnitschaja, A. Mazurakova, et al., "Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—risks and mitigation," Epma Journal, vol. 2023. Springer, 2023. springer.com

[8] K.Y. Lin, W.H. Hsih, Y.B. Lin, C.Y. Wen, et al., "Update in the epidemiology, risk factors, screening, and treatment of diabetic retinopathy," Journal of Diabetes Investigation, vol. 2021, Wiley Online Library, 2021. wiley.com

[9] N. Sikder, M. Masud, A. K. Bairagi, A. S. M. Arif, A. A. Nahid, "Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images," Symmetry, 2021. [Online]. Available: mdpi.com mdpi.com

[10] I. Odeh, M. Alkasassbeh, et al., "Diabetic retinopathy detection using ensemble machine learning," in Proc. ... Conference on Information ..., 2021. [PDF]

[11] G.T. Reddy, S. Bhattacharya, et al., "An ensemble-based machine learning model for diabetic retinopathy classification," in Proc. ... on emerging trends ..., 2020. [HTML]

[12] E. AbdelMaksoud, S. Barakat, and M. Elmogy, "A comprehensive diagnosis system for early signs and different diabetic retinopathy grades using fundus retinal images based on pathological changes detection," Computers in Biology and Medicine, vol. 2020, Elsevier, 2020. [HTML]

[13] S. Gayathri, V.P. Gopi, and P. Palanisamy, "A lightweight CNN for Diabetic Retinopathy classification from fundus images," in Biomedical Signal Processing and Control, vol. 55, 2020, Elsevier. [HTML]

[14] M. Akil, Y. Elloumi, and R. Kachouri, "Detection of retinal abnormalities in fundus image using CNN deep learning networks," in State of the Art in Neural Networks and their Applications, Elsevier, 2021. hal.science

[15] L. P. Cen, J. Ji, J. W. Lin, S. T. Ju, H. J. Lin, T. P. Li, et al., "Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks," Nature, 2021. nature.com

[16] Kannan, Rajeswari, S. R. Vispute, Reena Kharat, Dipti Salunkhe, and N. Vivekanandan. 2023. “Early Detection of Diabetic Retinopathy Using Deep Convolutional Neural Network”. Communications in Mathematics and Applications 14 (3):1283-92. https://doi.org/10.26713/cma.v14i3.2413.

[17] Sahlsten, J., Jaskari, J., Kivinen, J. et al. “Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading”. Sci Rep 9, 10750 (2019). https://doi.org/10.1038/s41598-019-47181-w

[18] Oh, K., Kang, H.M., Leem, D. et al. “Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images”. Sci Rep 11, 1897 (2021). https://doi.org/10.1038/s41598-021-81539-3

[19] Wejdan L. Alyoubi *, Wafaa M. Shalash, Maysoon F. Abulkhair Information Technology Department, University of King Abdul Aziz, Jeddah, Saudi Arabia https://doi.org/10.1016/j.imu.2020.100377

[20] Hamid Safi, MD, MPHa , Sare Safi, PhDb , Ali Hafezi-Moghadam, MD, PhDc , Hamid Ahmadieh, MDa, * a Ophthalmic Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran b Ophthalmic Epidemiology Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran c Molecular Biomarkers Nano-Imaging Laboratory, Brigham and Women's Hospital, Department of Radiology, Harvard Medical School, Boston, Massachusetts, USA https://www.scribd.com/document/526413601/safi2018

[21] S. Qummaret al., "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection," in IEEE Access, vol. 7, pp. 150530-150539, 2019, doi: 10.1109/ACCESS.2019.2 947484.

[22] Bilal A, Imran A, Baig TI, Liu X, Long H, et al. (2024) Improved Support Vector Machine based on CNN-SVD for vision-threatening diabetic retinopathy detection and classification. PLOS ONE 19(1): e0295951. https://doi.org/10.1371/journal.pone.0295951

[23] Michael D. Abràmoff MD, PhD Joseph M., Reinhardt PhD ,Stephen R. Russell MD, James C.Folk MD, Vinit B. Mahajan MD, PhD, Meindert Niemeijer PhD and Gwénolé Quellec PhD. Ophthalmology, 2010-06-01, Volume 117, Issue 6, Pages 1147-1154, Copyright © 2010 American Academy of Ophthalmology https://doi.org/10.1016/j.ophtha.2010.03.046

[24] Borys Tymchenko1 a , Philip Marchenko2 b and Dmitry Spodarets3 "Deep Learning Approach to Diabetic Retinopathy Detection" arXiv:2003.02261v1 [cs.LG] 3 Mar 2020 https://arxiv.org/

[25] Mishra, S., Hanchate, S., & Saquib, Z. (2020). Diabetic Retinopathy Detection using Deep Learning. 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE). https://doi.org/10.1109/icstcee49637.2020.9277506