**FMDB** *Transactions on Sustainable Health Science Letters*

# An Improved Model for Diabetic Retinopathy Detection by Using Transfer Learning and Ensemble Learning

**Muhammad Simul Hasan Talukder[1,*], Ajay Kirshno Sarkar[2], Sharmin Akter[3], Muhammad Nuhi-Alamin[4], Rejwan Bin Sulaiman[5]**

[1]Department of Atomic Energy Regulatory, Bangladesh Atomic Energy Regulatory Authority, Bangladesh.
[2,4]Department of Electrical and Electronic Engineering, Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh.
[3]Department of Biomedical Engineering, Jashore University of Science and Technology, Jashore, Bangladesh.
[5]Department of Artificial Intelligence and Cybersecurity, Northumbria University, North East of England, United Kingdom.
simulhasantalukder@gmail.com[1], aksarkar@eee.ruet.ac.bd[2], sharmintalukder120@gmail.com[3], nuhialamin@eee.ruet.ac.bd[4], Rejwan.sulaiman@northumbria.ac.uk[5]

**Abstract:** Diabetic Retinopathy (DR) is an ocular condition caused by a sustained high blood sugar level, which causes the retinal capillaries to block and bleed, causing retinal tissue damage. It usually results in blindness. Early detection can help in lowering the risk of DR and its severity. The robust and accurate prediction and detection of diabetic retinopathy is challenging. This paper develops a machine-learning model for detecting Diabetic Retinopathy that is entirely accurate. Pre-trained models such as ResNet50, InceptionV3, Xception, DenseNet121, VGG19, NASNetMobile, MobileNetV2, DensNet169, and DenseNet201 with pooling layer, dense layer, and appropriate dropout layer at the bottom of them were carried out in transfer learning (TL) approach. Data augmentation and regularization were performed to reduce overfitting. Transfer Learning model of DenseNet121, Average and weighted ensemble of DenseNet169 and DenseNet201 TL architectures contribute the highest accuracy of 100%, the highest precision, recall, F-1 score of 100%, 100%, and 100% individually.

**Keywords:** Diabetic Retinopathy; Transfer Learning; Ensemble Learning; Augmentation of Retinal Images; Introduction Regularization; Convolution Neural Network; Diabetes Mellitus; Diabetic Image of Fundus.

**Received on:** 27/12/2022, **Revised on:** 01/03/2023, **Accepted on:** 04/06/2023, **Published on:** 02/12/2023

**Cited by:** M. S. Hasan Talukder, A. Kirshno Sarkar, S. Akter, M. Nuhi-Alamin, and R. Bin Sulaiman, "An Improved Model for Diabetic Retinopathy Detection by Using Transfer Learning and Ensemble Learning," *FMDB Transactions on Sustainable Health Science Letters*, vol. 1, no. 2, pp. 92 –106, 2023.

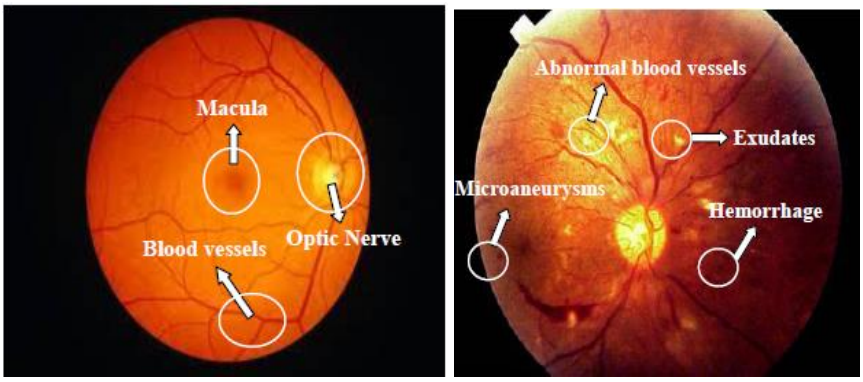


## 1. Introduction

Diabetes, commonly known as diabetes mellitus, is a condition in which the human body produces excess blood glucose [1]. It is a universal chronic disease that has been identified as the fourth leading cause of death. [2]. Diabetes has been related to several illnesses, including nerve damage, heart disease, stroke, foot difficulties, gum disease, and more [1]. Diabetes is anticipated to affect 336 million people globally, according to the International Diabetes Federation (IDF), with a 7.7% increase expected by 2030 [3, 4]. Diabetic Retinopathy (DR) is a diabetic condition in which the retinal blood vessels enlarge and spill fluid and blood [5]. According to the Mayo Clinic [6], frequent symptoms of DR include visual spots, colour impairment, blurred or fluctuating vision, and, in severe cases, complete vision loss in one or both eyes. Long-term high blood sugar levels cause blockage in the retina's micro-vessels, which are critical for nourishing the retina tissues. As a result, the eye strives to create new arteries to provide the retina with the nutrition and oxygen it requires; however, these newly formed vessels are

[*]Corresponding author.

weak and prone to blood loss, resulting in a retinal haemorrhage [7]. In both type 1 and type 2 diabetics, it is a significant cause of blindness [8]. Type-2 diabetes accounts for the majority of diabetes cases [9]. Figure 1 depicts the normal and DR-affected retinas, respectively.




(a) (b)

**Figure 1:** Fundus Image (a) Normal; (b) DR-affected [10]

Based on structural differences in colour fundus pictures, diabetic retinopathy is divided into Non-proliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR) [36]. Hard exudates, Microaneurysms, Soft exudates, and Hemorrhages, are some of the symptoms of NPDR, while Neovascularization and Vitreous haemorrhage are indicators of PDR [10]. However, early identification of diabetic retinopathy is crucial for preventing vision loss [37]. A quick clinical check and decision are frequently required for several reasons, such as a significant proportion of patients at a single institution or an essential and crucial patient condition. Screening a big group of people by hand is time-consuming and labour-intensive [38].

Furthermore, all patients should receive affordable treatment. Patients in many underdeveloped nations do not have access to quality health care or expensive treatment [39]. As a result of the absence of sufficient treatment, many indigent individuals are at risk of losing their sight [40].

Consequently, there is a need for reliable auto-DR screening methods using artificial intelligence to detect DR [41]. In this work, we have developed an efficient model for detecting DR via an ensemble of different transfer learning models. It has contributed to an excellent outcome [42]. The contribution of this article can be listed as follows:

- Proposing highly accurate transfer and ensemble models.
- Performance analysis of nine pre-trained models.
- Introduction regularization in each model.
- Replacement of a fully connected layer with a global average pooling layer
- Validating the average ensemble learning with weighted ensemble learning concept.
- Comparative analysis with the state of the existing work.

Section 3 represents the Materials and Methodology in the rest of the article. Section 4 shows the work results and discussion, and Section 5 contains the conclusion of the work.

## 2. Literature Review

Much deep learning-based research has been conducted on diabetic retinopathy (DR) detection from fundus images. This section discusses some of the existing research works. Zago et al. [11] employed the likelihood of lesion patches to categorize diabetic retinopathy or non-DR fundus pictures using two CNNs (pre-trained VGG16 and CNN). The DIARETDB1 dataset was utilized for training. The DDR, IDRiD, Messidor, DIARETDB0, Messidor-2, and Kaggle datasets were utilized for testing. The Messidor dataset delivered the best outcomes, with an AUC of 0.912 and a sensitivity of 0.94. A fundus image dataset can be classified as referable or non-referable DR using the model presented by Jiang et al. [12] using three CNNs (Inception-v3, ResNet152, and Inception-ResNet-v2). Before CNN training, the images were scaled, improved, and augmented, and the Adaboost approach was used to combine them. The network weights were updated using the Adam optimizer, and the system obtained an accuracy of 88.21 % and an AUC of 0.946. According to the DR severity levels, Pratt et al.[13] divided Kaggle fundus photos into five classes using CNN with ten convolutional layers, eight max-pooling layers, three fully connected layers, and a softmax classifier. Normalized and resized colour fundus images L2 regularization and dropout methods were used to reduce overfitting.

The model produced results with 95% specificity, 75% accuracy, and 30% sensitivity. Jayakumari et al. [14] proposed a transfer learning model where Inception V3 was used as a pre-trained model, and the dropout layer was used to avoid overfitting. The model had a training accuracy of 98.6% using the Kaggle dataset. The model's accuracy for no DR is 86.6 %, mild is 62.5 %, moderate is 66.6 %, severe is 57.1 %, and PDR is 42.8 %. Shaohua Wan et al. [15] adopted AlexNet, VggNet, GoogleNet, and ResNet with transfer learning and hyper-parameters tunning for analyzing diabetic image classification on the Kaggle dataset. VggNet-s by hyper-parameters contributed the best accuracy of 95.68. The severity of DR was classified by Narayana Bhagirath Thota et al. [16] using the VGG-16 model as a pre-trained neural network for fine-tuning. On high-quality photos, data augmentation, batch normalization, dropout layers, and learn-rate scheduling were used to obtain an accuracy of 74%.

Sabbir et al. [17] proposed an ensemble of SVM, KNN, and Naïve Bayes model, which was applied to the MESSIDOR fundus dataset. It achieved 92% accuracy. A deep learning model incorporating transfer learning from VGG16 was created by Islam et al. [18]. The new Kaggle dataset, "APTOS 2019 Blindness Detection," cut training time and produced an average accuracy of 0.9132683. CNN (VGGnet) was utilized by Habib Raj et al. [19] to estimate diabetic retinopathy (DR) and achieved 95.41% accuracy. Inception-ResNet-v2 was previously trained using transfer learning, and then a custom block of CNN layers was built on top of Inception-ResNet-v2 to create the hybrid model, according to Gangwar and Ravi's [20] proposal. On the Messidor-1 and APTOS datasets, the model has test accuracy of 72.33 % and 82.18 %, respectively.

Qummar et al. [21] trained an ensemble of five deep Convolution Neural Network (CNN) models (Resnet50, Xception, Inceptionv3, Dense169, and Dense121) using the publicly accessible Kaggle dataset of retina images and reached an accuracy of 80.70 %. In order to enhance image quality and consistently equalize intensities, Momeni Pour et al. [22] created a new diabetic retinopathy monitoring model that used the contrast-limited Adaptive Histogram Equalization approach. The EffcientNet-B5 architecture is then used for the classification step. This network's effectiveness lies in consistently scaling all its dimensions. The final model is trained using a blend of the Messidor-2 and IDRiD datasets and then validated on the Messidor dataset. The area under the curve (AUC) is raised to 0.945 from 0.936, the maximum value in all recent works. The convolutional Block Attention Module (CBAM) was built on top of the encoder by Farag et al. [23] to increase its discriminative power. They used the encoder from DenseNet169 to generate a visual embedding. Applying the APTOS dataset, the model contributed 97% accuracy. A summary of the review work is presented in Table 1.

**Table 1:** Summary of related work

| Ref. | Model | Dataset | No. of Image | Augmentation | Transfer Learning | Ensemble Learning | Accuracy (%) |
|---|---|---|---|---|---|---|---|
| [17] | SVM+ KNN+Naïve Bayes | Messidor dataset | 1200 | No | No | yes | 92 |
| [18] | VGG16 | APTOS 2019 Blindness Detection | 5590 | No | yes | No | 80 |
| [19] | VGGnet | Kaggle | 35126 | No | No | No | 95.41 |
| [24] | VGG-NiN | EyePACS | 88,702 | No | No | No | 85 |
| [20] | Inception-ResNet-v2 | Messidor-1 | 1200 | yes | yes | No | 72.33 |
| | | APTOS | 5590 | | | | 82.81 |
| [21] | Resnet50, Inceptionv3, Xception, DenseNet121, DenseNet169 | Kaggle | 35126 | yes | No | Yes | 80.7 |
| [22] | EfficientNet-B5 | Mixture of Messidor-2 and IDRiD | 1748 | No | No | No | 94.5 |
| | | | 516 | | | | |
| [23] | DenseNet169 and Convolutional Block Attention Module | APTOS | 5590 | No | yes | No | 97 |

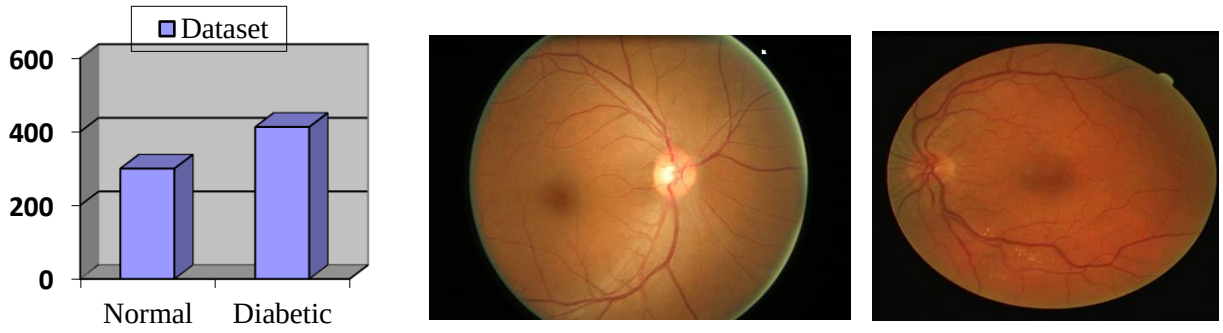
From the literature, it is obvious that initially, the researcher used traditional ML methods in DR detection. Day by day, the CNN transfer learning approach was becoming popular [43]. Regularization, replacement of fully connected layer by global average pooling layer, updated pre-trained models, and ensemble learning concepts are not used in diabetic retinopathy detection [44]. The performance of those studies was also not so high. We have included and resolved the issues in this study, and a comprehensive analysis has been carried out [45].

## 3. Materials and Methodology

In order to diagnose diabetic retinopathy from fundus images, this research offers nine transfer learning models. A combination of two publicly available datasets has been used to carry out these experiments. The following section describes the whole methods and experimental setup in detail.

### 3.1. Dataset

For this experiment, we combined the Diseases Grading of Indian Diabetic Retinopathy Image Dataset (IDRID) [25] with the fundus dataset from Mendeley [26]. One thousand five hundred genuine colour fundus photos in 24-bit RGB format are divided into 26 categories in the dataset [26]. Diseases Grading of IDRID consists of original colour fundus images (516 images divided into train set (413 images) and test set (103 images) [25]. We have taken 300 normal fundus images from the Mendeley dataset and a trainset of 431 diabetics' fundus images from the IDRID dataset. So, our dataset consists of two classes named “Normal” and “Diabetic”. Figure 2 displays the distribution of images by ease class. Our dataset is quite small compared to the Messidor dataset [27]. However, the main aim of our work is to design a robust and accurate model with limited label data.




(a) (b) (c)

**Figure 2:** (a) Distribution of Data in each class; (b) Normal Image of Fundus; (c) Diabetic Image of Fundus

### 3.2. Data Augmentation

Deep Convolutional Neural Networks (DCNNs) have displayed impressive performance on numerous Computer Vision projects. However, these systems depend intensely on enormous datasets to dodge overfitting [28]. When a network learns a function with significant variance, the behaviour is overfitting. It occurs due to a lack of training dataset and insufficient diversity of training data or uneven class balance in the dataset [29]. Augmentation is a way to deal with such problems. The performance of DCNN will be enhanced with increasing training datasets with different augmentation techniques [30]. The diabetic retinopathy dataset was split into training, validation, and testing data folders using split folders.ratio() module in Python at ratio 0.70, 0.15, and 0.15, respectively, and resized at 224*224 pixels. Table 2 contains the summary of the datasets. Data augmentation techniques such as rotation, zooming, flipping, and shifting were applied in the training dataset in our work by using the ImageDatagenerator module of Python [31]. The parameters of augmentation are listed in Table 3.

**Table 2:** Training, Validation, and testing data distribution

| SI. No. | Dataset | No. of Images | | |
|---|---|---|---|---|
| | | Normal | Diabetic | Total |
| 1 | Training | 210 | 289 | 499 |

| | | | | |
|---|---|---|---|---|
| 2 | Validation | 45 | 61 | 106 |
| 3 | Testing | 45 | 63 | 108 |

**Table 3:** Data augmentation Techniques with parameters

| SI. No. | Augmentation Techniques | Parameters |
|---|---|---|
| 1 | Rotation | rotation_range = 10 |
| 2 | Zoom | zoom_range = 0.2 |
| 3 | Width shift | width_shift_range = 0.2 |
| 4 | Height shift | height_shift_range = 0.2 |
| 5 | Vertical flip | vertical_flip = True |
| 6 | Horizontal flip | horizontal_flip = True |

### 3.3. Transfer Learning

Using a learning strategy created for one assignment as the basis for a model on another assignment is called transfer learning (TL), a machine learning methodology [32]. It reuses the pre-trained model on new problems. The main benefits of TL are reduced training time, improved neural network performance (in most cases), and no need for a large amount of data [33]. The most common pre-trained models for TL are VGG19, VGG16, AlexNet, Inceptions, etc. [32]. In this work, ResNet50, InceptionV3, Xception, MobileNetV2, NASNetMobile, VGG19, DenseNet121, DenseNet169 and DenseNet201 were applied as pre-trained model [46]. The classification layer of each model is replaced by the GlobalAveragePooling2D layer, SoftMax layer, Dense layer with the number of 2 classes, and Dropout layer. Dropout value 0.25 was carried out to avoid overfitting [47]. All the models were fitted using the Adam optimizer with a learning rate of 0.001, categorical cross-entropy, and batch size 16. Table 4 shows the trainable parameters of the models [48]. Among all the pre-trained models, DenseNet121 provided the best outcome with 100% accuracy, and DenseNet169 and DenseNet201 performed pretty well and were chosen for ensemble learning [49].

**Table 4:** TL architectures applied in our works

| Model | Total Parameter | Trainable Parameters | Non-Trainable Parameters |
|---|---|---|---|
| InceptionV3 | 21,806,882 | 4,098 | 21,802,784 |
| Xception | 20,865,578 | 4,098 | 20,861,480 |
| DenseNet121 | 18,321,984 | 2,050 | 7,037,504 |
| DenseNet201 | 18,325,826 | 3,842 | 18,321,984 |
| DenseNet169 | 12,646,210 | 3,330 | 12,642,880 |
| ResNet50 | 23,591,810 | 4,098 | 23,587,712 |
| NASNetMobile | 4,271,830 | 2,114 | 4,269,716 |
| VGG19 | 20,025,410 | 1,026 | 20,024,384 |
| MobileNetV2 | 2,260,546 | 2,562 | 2,257,984 |

### 3.4. Ensemble Learning

Ensemble learning attempts to outperform any single algorithm by integrating several algorithms and combining the results with various voting processes [34]. It is used to decrease variance bias and improve the prediction. [21]. From the performance Table 5 and confusion matrix analysis, TL DenseNet169 and DenseNet201 architectures made predictions very well [50]. So, the average and weighted ensembles were performed separately on the models. Average ensemble and weighted ensemble are shown in algorithms 1 and 2.

**Algorithm 1. Average Ensemble of the Models**

**Input:** Test_set $T$: Models $Z_j$ (j $=$ 1 to m) where j is the number of models

**Output:** $I_o$

Ensemble_model B= [$Z_1$,$Z_2$, . . .$Z_j$ ]
**For** n = 1 to k **do**
  Predict, $P$ = generate ($T$)
  $Q$= add ($P$, along y axis)
  $I_o$= index_max (Q, along x axis)
Confusion_matrix ($I_o$, $T$)
Classification_ matrices ($I_o$, $T$)
**End**

**Algorithm 2. Weighted ensemble of the models**

**Input:** Test_set $T$: Models $Z_j$ and Weight_set $W_j$ (j = 1 to m) where j is the number of models.
**Output:** $I_o$
Ensemble_model B= [$Z_1$,$Z_2$, . . .$Z_j$ ]
**For** n = 1 to k **do**
  Predict, $P$ = generate ($T$)
  $Q = add\ (P * W_i, along\ y\ axis)$
  $I_o$= index_max (Q, along x axis)
Confusion_matrix ($I_o$, $T$)
Classification_ matrices ($I_o$, $T$)
**End**

In our work, E= [DenseNet169, DenseNet201] and Weight $W_k$ = [ 0.2 ,02]. K=1,2.

### 3.5. Proposed Methodology

This work represents a TL and Ensemble model in detecting diabetic retinopathy, as shown in Figure 3. The IDRID dataset from Kaggle and Fundus-Dataset from Mendeley was combined and split into training, validation, and testing with a ratio of 70%, 15%, and 15%, respectively [51]. The training dataset was augmented with parameter tuning using the techniques described in the previous section [52]. The nine pre-trained models in Table 4. were carried out in this experiment, adding the Average Global Pooling Layer, Dropout layer & dense layer at the bottom of base models [53]. The Adam optimizer and categorical cross-entropy loss function were used to train the networks precisely. The following equation (1) can express the categorical entropy loss [54].

$$loss = -\sum_{i=1}^{Output\ size} y_i * log_{\hat{y}_i} \quad (1)$$

Where $\hat{y}_i$ is the $i-th$ scalar value in the model output, $y_i$ the corresponding target value is the output size, which is the number of scalar values in the model output.

The models were trained, validated, and tested. Out of them, the best model was found out. Comparatively, two poor little models were ensembled to generate a new classifier. All the classifiers were evaluated using performance measures such as a confusion matrix, precision, recall, F1-score, and accuracy to choose the optimal model for Diabetic retinopathy diagnosis.

### 3.6. Performance Evaluation

Our proposed model's performance was evaluated both qualitatively and visually. The qualitative evaluation of image classification is widely used [35]. In addition, we quantified our model by measuring the parameters of Accuracy (Acc), Precision, recall, and F1-score.

$$\text{Accuracy} = \frac{\sum TP+\sum TN}{\sum TP+\sum TN+\sum FP+\sum FN} * 100 \quad (2)$$

$$\text{Precision} = \frac{\sum TP}{\sum TP+\sum FP} * 100 \quad (3)$$

$$\text{Recall} = \frac{\sum TP}{\sum TP+\sum FN} \quad (4)$$

$$\text{F1 score} = 2 * \left(\frac{Precision*Recall}{Precision+Recall}\right) * 100 \quad (5)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (7)$$

$$\text{Macro Avg Measure} = \frac{1}{N}(Measure\ in\ class_1 + Measure\ in\ class_2 + \cdots + Mesure\ in\ class_N) \quad (8)$$

$$\text{Weighted Average Measure} = \frac{1}{Total\ number\ of\ sample}\ [(Measure * weight)\ in\ Class_1 + (Measure * weight) in\ class_2 + \cdots + (Measure * weight) in\ class_N] \quad (9)$$

Where: TP stands for True Positive, TN denotes True Negative, FP is False Positive, and FN is False Negative.

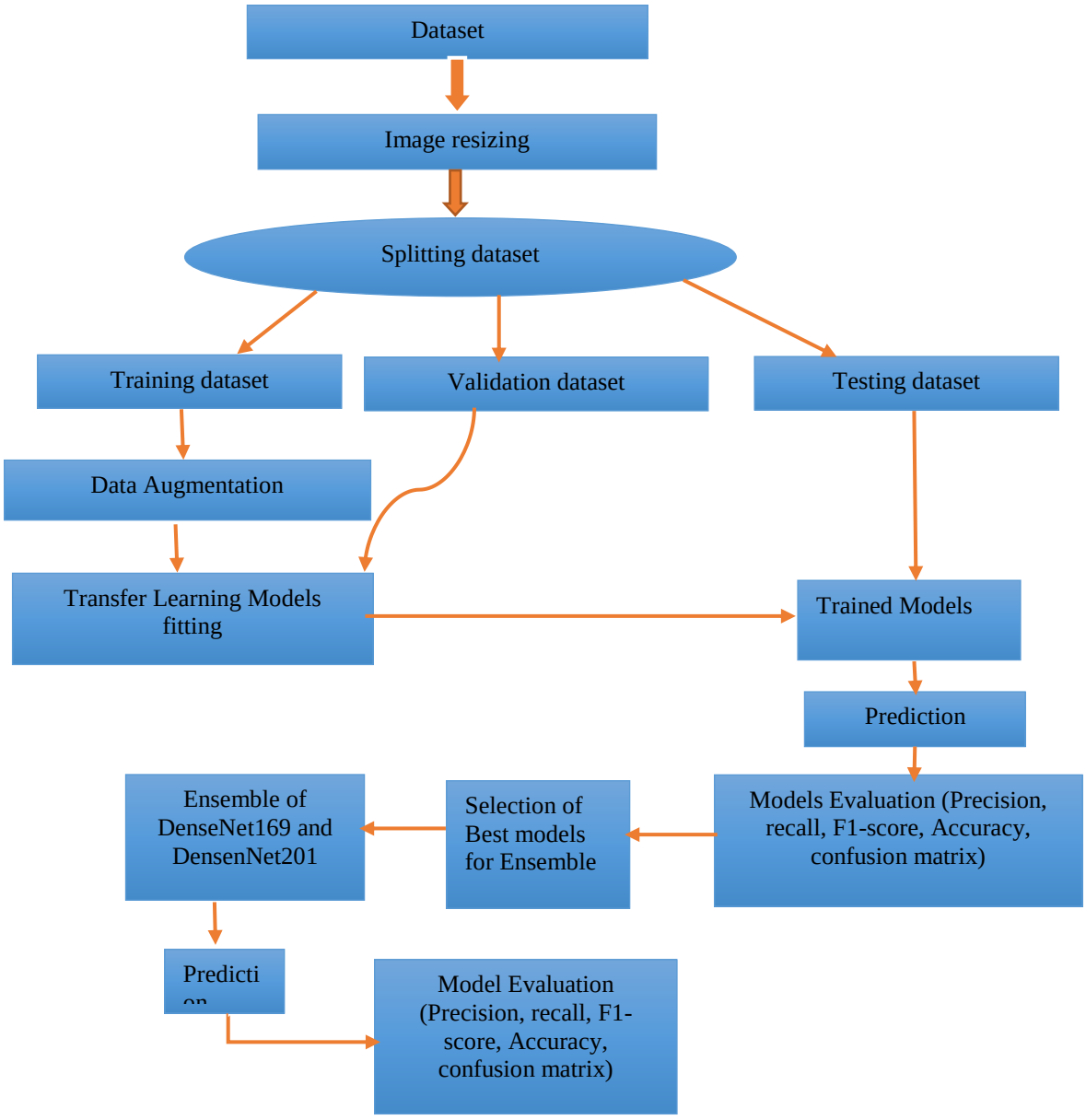



**Figure 3:** Proposed Methodology

## 4. Result and Discussion

In this work, Different transfer learning architectures are implemented to detect diabetic retinopathy. It was carried out on GoogleColab with GPU. Different pre-trained models such as DesnseNet121, DenseNet201, DenseNet169, InceptionV3, Xception, ResNet50, NASNetMobile, VGG19, and MobileNetV2 were used with adding Global Average Pooling layer, dropout layer and Dense layer at the beneath of base model. Data augmentation with hyperparameter tuning in Table 3 was applied to the training dataset. The Training loss and validation loss are presented in Figure 4. In the case of inceptionV3, Xecption, and ResNet50, the Validation loss is smaller than the training loss. Overfitting exists here, and Underfitting is in VGG19 since validation loss is greater than training loss.

The rest are reached almost at optimal fit. DenseNet201, DenseNet169, and DenseNet121 are excellent, and DenseNet121 is the best in optimal model fit. Similarly, in Figure 5. In the case of InceptionV3, Xecption, ResNet50, and VGG19, the model fit in training and validation accuracy is quite poor. The rest are pretty good, but the DenseNet121, DenseNet201, and DenseNet169 are superior. Figure 6. represents the confusion matrix for each TL and Ensemble Learning model, where the number of classifications and misclassifications are easily visible. ResNet50 shows the highest misclassification number of 22. DenseNet121 provides the best prediction where all the testing data is properly classified. The average and weighted ensemble with the same weight of DenseNet201 and DenseNet169 also results from the exact classification. Precision, recall, f1 score, and accuracy have been calculated for individual classes in Table 5. DenseNet121, DenseNet169, DenseNet201, InceptionV3 NASNetMobile, VGG19, and ensemble models show 100% precision, but NASNetMobile and VGG19 show lower recall and F1-score on the 'Normal' class dataset. DesneNet121, ensemble models Xception and MobileNetV2 outperform the rest in precision, but Xception and MobileNetV2 result in lower recall and F1 scores. From the analysis, DenseNet121 and Ensemble models provide 100% precision, recall, and F1-score on both classes in Table 5 and also remain the same in overall performance in Table 6. So, DenseNet121 and Ensemble models are taken as benchmarks for DR detection.

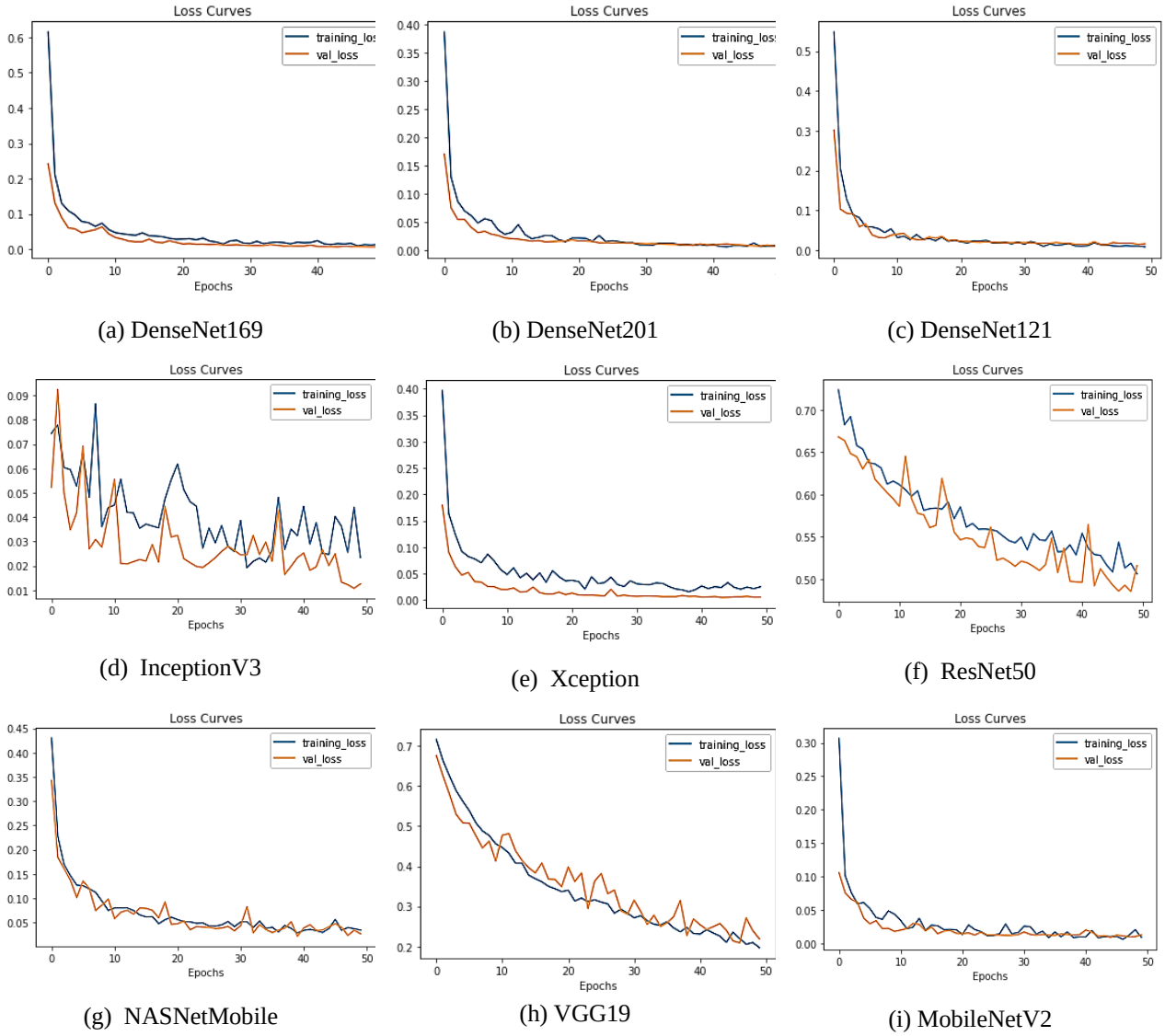
**Table 5:** Performance Analysis of each model in each class

| Base Model | Normal | | | Diabetic | | | Accuracy (%) |
|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1- score | Precision | Recall | F1- score | |
| Dense Net169 (TL) | 100 | 98 | 99 | 98 | 100 | 99 | 99.07 |
| DenseNet201 (TL) | 100 | 98 | 99 | 98 | 100 | 99 | 99.07 |
| InceptionV3 | 100 | 98 | 99 | 98 | 100 | 99 | 99.07 |
| Xception | 98 | 100 | 99 | 100 | 98 | 99 | 99.07 |
| ResNet50 | 91 | 69 | 79 | 71 | 92 | 80 | 79.62 |
| NASNetMobile | 100 | 96 | 98 | 97 | 100 | 98 | 98.14 |
| VGG19 | 100 | 92 | 96 | 94 | 100 | 97 | 96.29 |
| MobileNetV2 | 96 | 100 | 98 | 100 | 97 | 98 | 98.14 |
| DenseNet121 (TL) | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Average Ensemble of DenseNet169 (TL) and DeseNet201 (TL) | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Weighted Ensemble of DenseNet169 (TL) and DeseNet201 9 (TL) | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

**Table 6:** Overall performance Analysis of each model

| Base Model | Macro average | | | Weighted average | | | Accuracy (%) |
|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1-score | Precision | Recall | F1-score | |
| DenseNet169(TL) | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| DenseNet201(TL) | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| InceptionV3 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| Xception | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| ResNet50 | 81 | 81 | 80 | 82 | 80 | 80 | 80 |
| NASNetMobile | 98 | 98 | 98 | 98 | 98 | 98 | 98 |
| VGG19 | 97 | 96 | 96 | 97 | 96 | 96 | 96 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MobileNetV2 | 98 | 98 | 98 | 98 | 98 | 98 | 98 |
| DenseNet121 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Average Ensemble of DenseNet169 (TL) and DeseNet201 (TL) | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Weighted Ensemble of DenseNet169 (TL) and DeseNet201 9 (TL) | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

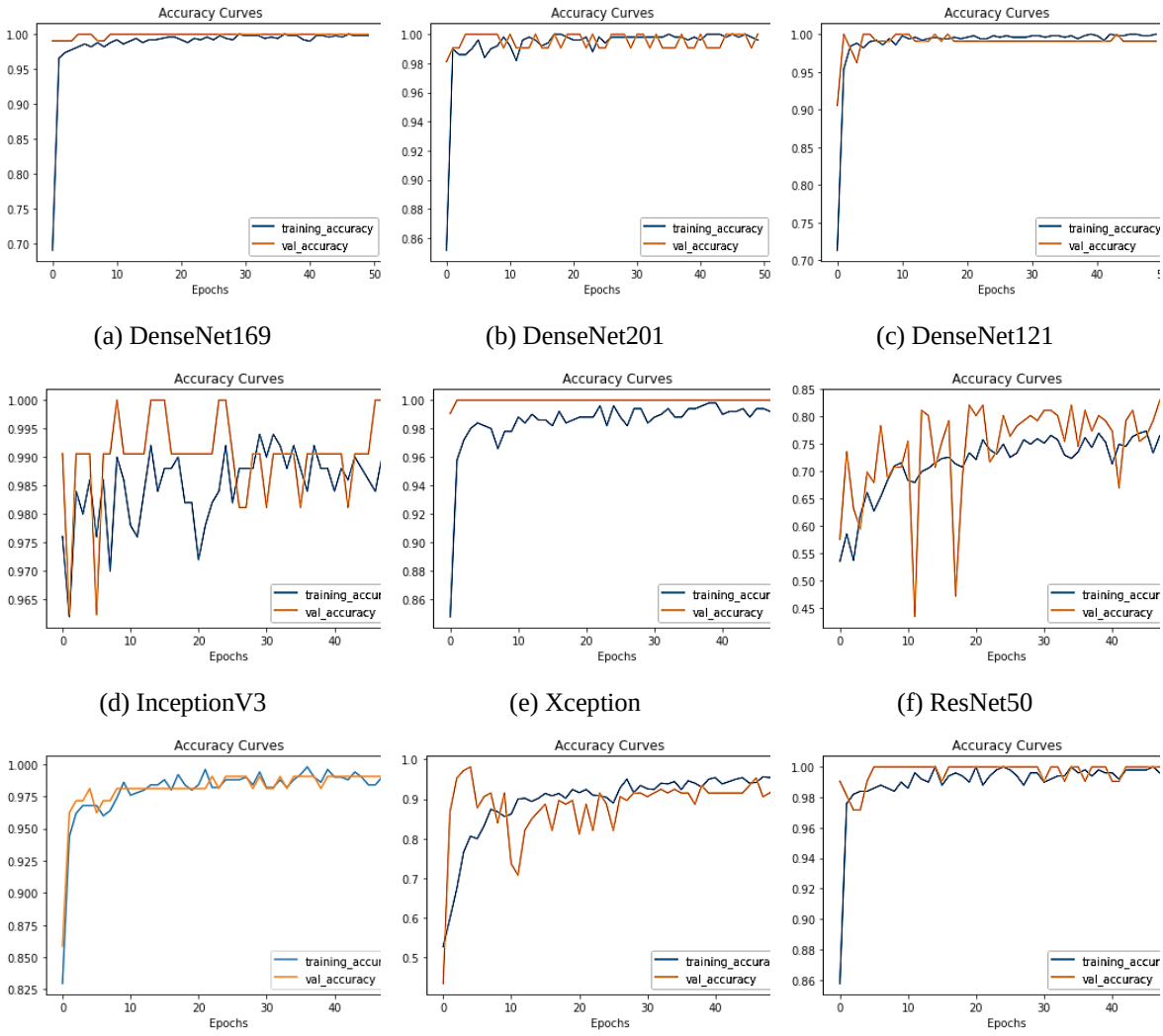



(a) DenseNet169 (b) DenseNet201 (c) DenseNet121

(d) InceptionV3 (e) Xception (f) ResNet50

(g) NASNetMobile (h) VGG19 (i) MobileNetV2

**Figure 4:** Training loss and Validation Loss when the base model (a) DenseNet169; (b)DenseNet201; (c) DenseNet121; (d) InceptionV3; (e) Xception; (f) ResNet50; (g) NASNetMobile; (h) VGG19; (i) MobileNetV2

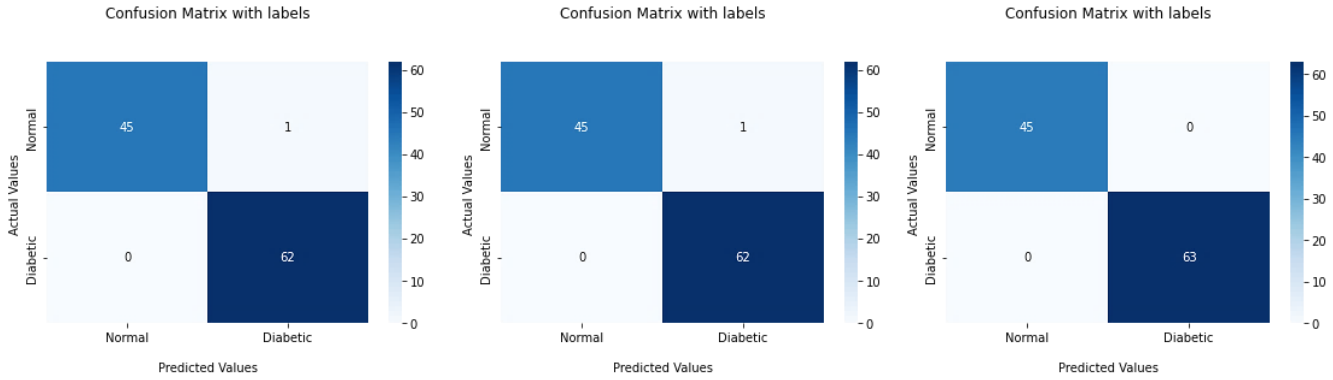



(a) DenseNet169 (b) DenseNet201 (c) DenseNet121

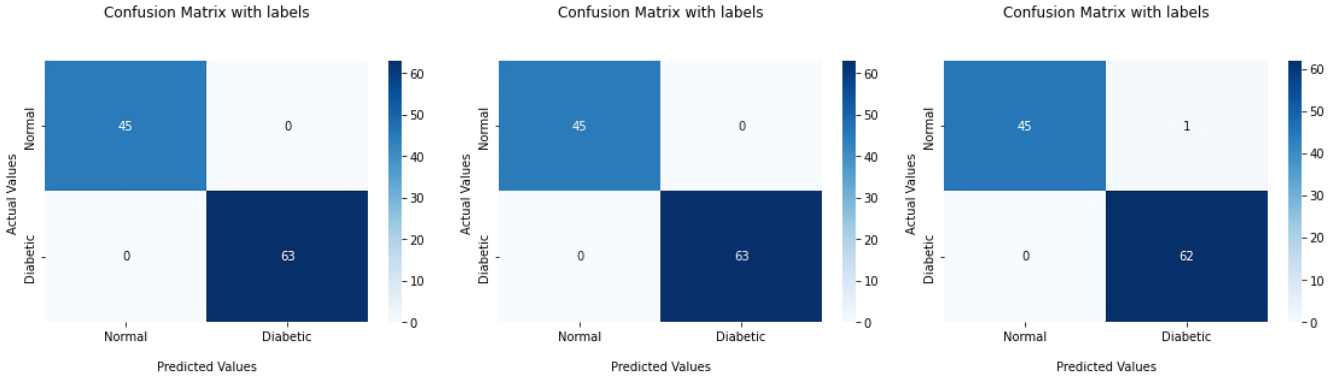
(d) InceptionV3 (e) Xception (f) ResNet50

(g) NASNetMobile (e) VGG19 (i) MobileMetV2

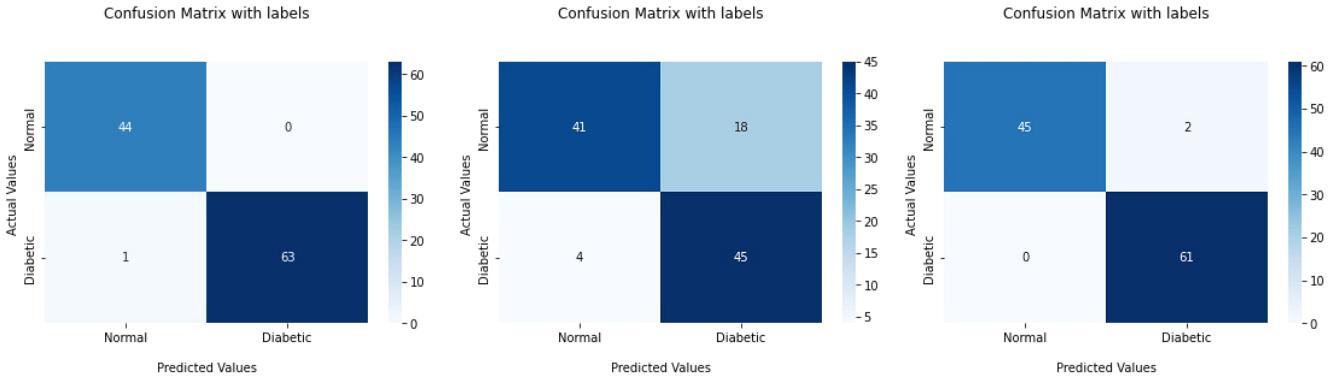
**Figure 5:** Training accuracy and Validation accuracy when the base model (a) DenseNet169; (b)DenseNet201; (c) DenseNet121; (d) InceptionV3; (e) Xception; (f) ResNet50; (g) NASNetMobile; (h) VGG19; (i) MobileNetV2



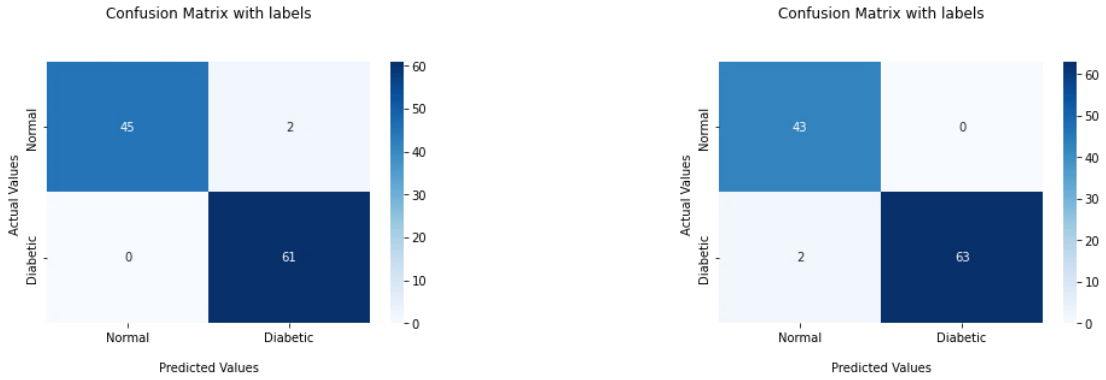

(a) DenseNet169 (b) DenseNet 201 (c) Average Ensemble of DenseNet169 and 201




(d) Weighted Ensemble of DenseNet169 and 201 (e) DenseNet121 (f) InceptionV3




(g) Xception (h) ResNet50 (i) NASNetMobile




(j) VGG19 (k) MobileNetV2

**Figure 6:** Confusion Matrix. (a) DenseNet169; (b) DenseNet201; (c) Average Ensemble of DenseNet169 and 201; (d) Weighted Ensemble of DenseNet169 and 201; (e) DenseNet121; (f) InceptionV3; (g) Xception; (h) ResNet50; (i) NASNetMobile; (j) VGG19; (k) MobileNetV2

Figure 7 represents the comparison of our models in testing accuracy. Our proposed DenseNet169 TL learning model outperforms the other models. A comparison with the existing work is also shown in Table 7.

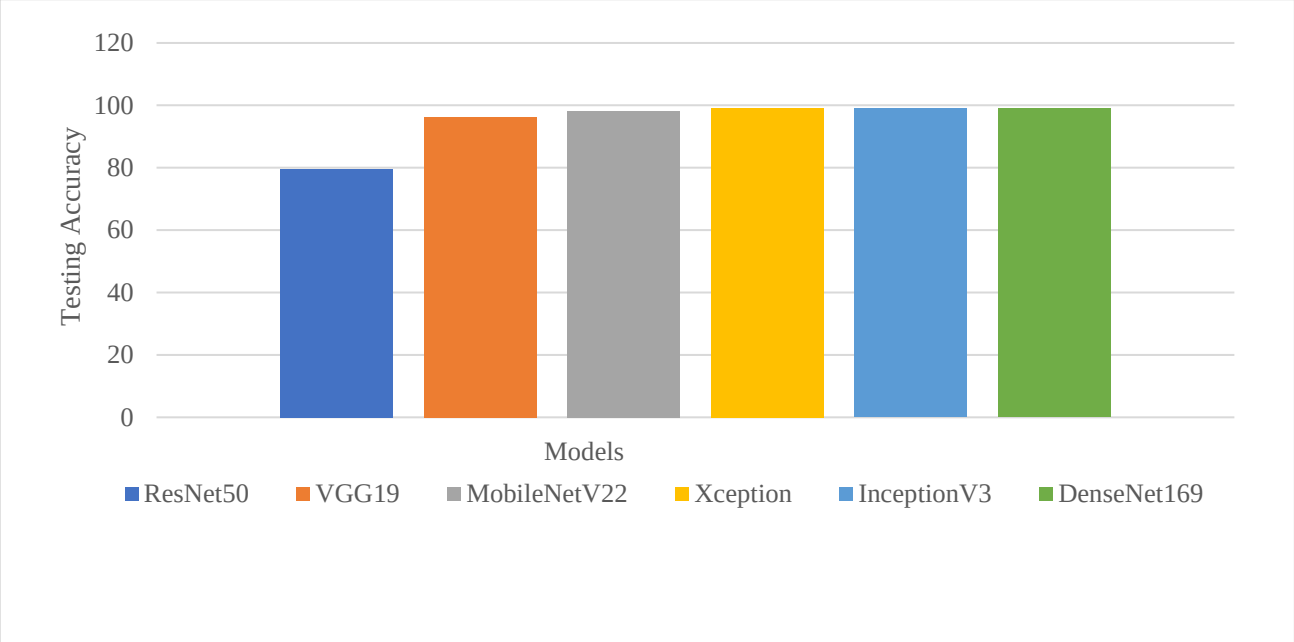



**Figure 7:** Comparison of testing accuracy of our models

**Table 7:** Comparison of our work with very recent work

| Models | Sensitivity (%) | Specificity (%) | Accuracy (%) |
|---|---|---|---|
| Mohamed M. Farag et al. [23] | 98.3 | 94.55 | 97 |
| Our proposed Model (DenseNet121) | 100 | 100 | 100 |
| Our Average Ensemble proposed model (DenseNet201 and DenseNet169) | 100 | 100 | 100 |
| Our Weighted Ensemble proposed model (DenseNet201 and DenseNet169) | 100 | 100 | 100 |

## 5. Conclusion

Automated screening systems significantly reduce the time required to determine diagnoses, saving ophthalmologists time and money and allowing patients to be treated more quickly. Automated DR detection systems play an important role in detecting DR at an early stage. In our work, DenseNet121 architecture provides the highest accuracy of 100% out of the individual TL architectures. The ensemble of DenseNet169 and DesneNet201 TL architectures also results from 100% accuracy, 100% sensitivity, and specificity. Data augmentation, parameter tuning, the Global Average Pooling layer, and the dropout layer at the bottom of the pre-trained model have played a critical role in our work. An accurate determination of diabetic retinopathy at an appropriate time may help the patients to take preventive action from the very beginning. The research has some limitations. Firstly, no conventional ML classifier is used since deep learning classifiers show superiority in image classification. Secondly, Data pre-processing has been ignored. But it is an important step of ML. Thirdly, only the existence of DR or not has been considered here. The severity level and other symptoms of ophthalmological diseases are not considered. In the future, this work will be carried out by considering all the limitations and will be tested with real-world data in the field.

**Acknowledgement:** MSH Talukder worked on implementation, result analysis, and methodology writing. Ak. Sarkar organized the conceptualization and drafting of the manuscript. Shramin Akter researched literature review and introduction writing. Finally, Nuhi-Alamin has reviewed and corrected the manuscript.

**Data Availability Statement:** The dataset is public.

**Funding Statement:** This research received no external funding.

**Conflicts of Interest Statement:** The authors declare that they have no conflict of interest.

**Ethics and Consent Statement:** The consent was obtained from the colleges during data collection, and ethical approval and participant consent were received.

## References

1. M. Farooq et al., "Untangling Computer-Aided Diagnostic System for Screening Diabetic Retinopathy Based on Deep Learning Techniques", Sensors, vol. 22, no. 5, p. 1803, 2022. Available: 10.3390/s22051803.
2. I. Qureshi, J. Ma and Q. Abbas, "Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning", 2022.
3. D. S. W. Ting, G. C. M. Cheung, and T. Y. Wong, “Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review,” Clin. Experiment. Ophthalmol., vol. 44, no. 4, pp. 260–277, 2016.
4. R. E. Washington, T. J. Orchard, V. C. Arena, R. E. LaPorte, A. M. Secrest, and E. S. Tull, “All-cause mortality in a population-based type 1 diabetes cohort in the U.S. Virgin Islands,” Diabetes Res. Clin. Pract., vol. 103, no. 3, pp. 504–509, 2014.
5. American academy of ophthalmology-what is diabetic retinopathy? [Online]. Available, https://www.aao.org/eye-health/diseases/what-is-diabetic-retinopath, [Accessed by Jan. 03, 2021].
6. “Diabetic retinopathy,” May 30, 2018. https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611, [Accessed by Jan. 03, 2021].
7. Odeh, M. Alkasassbeh and M. Alauthman, "Diabetic Retinopathy Detection using Ensemble Machine Learning," 2021 International Conference on Information Technology (ICIT)p. 173-178, 2021, doi: 10.1109/ICIT52682.2021.9491645.
8. S. Das and C. Malathy, “Survey on the diagnosis of diseases from retinal images,” J. of Physics: Conference Series, National Conf. on Mathematical Techniques and its Applications, 2018, pp. 1-11.
9. S. Atre, “Addressing policy needs for prevention and control of type 2 diabetes in India,” Perspect. Public Health, vol. 135, no. 5, pp. 257–263, 2015.
10. V. S and V. R, "A Survey on Diabetic Retinopathy Disease Detection and Classification using Deep Learning Techniques," 2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII), pp. 1-4, 2021, doi: 10.1109/ICBSII51839.2021.9445163.
11. Zago. GT, Andreao. RV, Dorizz. B and Teatini Salles. EO, “Diabetic retinopathy detection using red lesion localization and convolutional neural networks. Computers in Biology and Medicine”, 116, 103537, 2020.
12. Jiang. H, Yang. K, Gao. M, Zhang. D, Ma. H and Qian. W, “An interpretable ensemble deep learning model for diabetic retinopathy disease classification”, 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2045–2048, 2019.
13. Pratt. H, Coenen. F, Broadbent. DM, Harding. SP and Zheng. Y, “Convolutional neural networks for diabetic retinopathy”, Procedia Comput Sci., 90:200–5, 2016.
14. C. Jayakumari, V. Lavanya and E. P. Sumesh, "Automated Diabetic Retinopathy Detection and Classification using ImageNet Convolution Neural Network using Fundus Images," 2020 International Conference on Smart Electronics and Communication (ICOSEC), pp. 577-582, 2020, doi: 10.1109/ICOSEC49089.2020.9215270.
15. S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," 2022.
16. N. B. Thota and D. Umma Reddy, "Improving the Accuracy of Diabetic Retinopathy Severity Classification with Transfer Learning," 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 1003-1006, 2020, doi: 10.1109/MWSCAS48704.2020.9184473.
17. M. M. H. Sabbir, A. Sayeed and M. A. -U. -Z. Jamee, "Diabetic Retinopathy Detection using Texture Features and Ensemble Learning," 2020 IEEE Region 10 Symposium (TENSYMP), pp. 178-181, 2020, doi: 10.1109/TENSYMP50017.2020.9230600.
18. M. R. Islam, M. A. M. Hasan and A. Sayeed, "Transfer Learning based Diabetic Retinopathy Detection with a Novel Pre-processed Layer," 2020 IEEE Region 10 Symposium (TENSYMP), pp. 888-891, 2020, doi: 10.1109/TENSYMP50017.2020.9230648.
19. M. A. Habib Raj, M. A. Mamun, and M. F. Faruk, "CNN Based Diabetic Retinopathy Status Prediction Using Fundus Images," 2020 IEEE Region 10 Symposium (TENSYMP), pp. 190-193, 2020, doi: 10.1109/TENSYMP50017.2020.9230974.

20. A. Gangwar and V. Ravi, "Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning," Evolution in Computational Intelligence, pp. 679-689, 2020. Available: 10.1007/978-981-15-5788-0_64.
21. S. Qummar et al., "A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection," in IEEE Access, vol. 7, pp. 150530-150539, 2019, doi: 10.1109/ACCESS.2019.2947484.
22. A. Momeni Pour, H. Seyedarabi, S. H. Abbasi Jahromi and A. Javadzadeh, "Automatic Detection and Monitoring of Diabetic Retinopathy Using Efficient Convolutional Neural Networks and Contrast Limited Adaptive Histogram Equalization," in IEEE Access, vol. 8, pp. 136668-136673, 2020, doi: 10.1109/ACCESS.2020.3005044.
23. M. M. Farag, M. Fouad, and A. T. Abdel-Hamid, "Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module," in IEEE Access, vol. 10, pp. 38299-38308, 2022, doi: 10.1109/ACCESS.2022.3165193.
24. Z. Khan et al., "Diabetic Retinopathy Detection Using VGG-NIN a Deep Learning Architecture," in IEEE Access, vol. 9, pp. 61408-61416, 2021, doi: 10.1109/ACCESS.2021.3074422.
25. Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, Fabrice Meriaudeau, April 24, 2018, "Indian Diabetic Retinopathy Image Dataset (IDRiD)", IEEE Dataport, doi https://dx.doi.org/10.21227/H25W98.
26. R. Saeed, “fundus-dataset.zip.” figshare, figshare. Dataset. https://doi.org/10.6084/m9.figshare.16986166.v1 11-Nov-2021 [Accessed by 11 Nov 2021].
27. Decencière et al. Feedback on a publicly distributed database: the Messidor database.Image Analysis & Stereology, v. 33, n. 3, p. 231-234, 2014.
28. C. Shorten and T. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning", Journal of Big Data, vol. 6, no. 1, 2019. Available: 10.1186/s40537-019-0197-0.
29. A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," 2018 International Interdisciplinary PhD Workshop (IIPhDW), pp. 117-122, 2018, doi: 10.1109/IIPHDW.2018.8388338.
30. J. Shijie, W. Ping, J. Peiyi and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," 2017 Chinese Automation Congress (CAC), pp. 4165-4170, 2017, doi: 10.1109/CAC.2017.8243510.
31. P. Bansal, R. Kumar, and S. Kumar, "Disease Detection in Apple Leaves Using Deep Convolutional Neural Network", Agriculture, vol. 11, no. 7, p. 617, 2021.
32. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. IEEE Trans. Knowl. Data Eng., 22, 1345–1359, 2010.
33. What Is Transfer Learning? Exploring the Popular Deep Learning Approach. Available at: https://builtin.com/data-science/transfer-learning, [Accessed by 11 Nov 2023].
34. M. Sharma, K. Nath, R. Sharma, C. Kumar, and A. Chaudhary, "Ensemble Averaging of Transfer Learning Models for Identification of Nutritional Deficiency in Rice Plant", Electronics, vol. 11, no. 1, p. 148, 2022. Available: 10.3390/electronics11010148.
35. T. Tran, J. Choi, T. Le, and J. Kim, "A Comparative Study of Deep CNN in Forecasting and Classifying the Macronutrient Deficiencies on Development of Tomato Plant", Applied Sciences, vol. 9, no. 8, p. 1601, 2019. Available: 10.3390/app9081601.
36. R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automatic Detection of Diabetic Eye Disease Through Deep Learning Using Fundus Images: A Survey," in IEEE Access, vol. 8, pp. 151133-151149, 2020, doi: 10.1109/ACCESS.2020.3015258.
37. T. Liu et al., "A Novel Diabetic Retinopathy Detection Approach Based on Deep Symmetric Convolutional Neural Network," in IEEE Access, vol. 9, pp. 160552-160558, 2021, doi: 10.1109/ACCESS.2021.3131630.
38. M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers," in IEEE Access, vol. 8, pp. 76516-76531, 2020, doi: 10.1109/ACCESS.2020.2989857.
39. M. Mateen, J. Wen, M. Hassan, N. Nasrullah, S. Sun and S. Hayat, "Automatic Detection of Diabetic Retinopathy: A Review on Datasets, Methods and Evaluation Metrics," in IEEE Access, vol. 8, pp. 48784-48811, 2020, doi: 10.1109/ACCESS.2020.2980055.
40. F. Saeed, M. Hussain and H. A. Aboalsamh, "Automatic Diabetic Retinopathy Diagnosis Using Adaptive Fine-Tuned Convolutional Neural Network," in IEEE Access, vol. 9, pp. 41344-41359, 2021, doi: 10.1109/ACCESS.2021.3065273.
41. V. S and V. R, "A Survey on Diabetic Retinopathy Disease Detection and Classification using Deep Learning Techniques," 2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII), pp. 1-4, 2021, doi: 10.1109/ICBSII51839.2021.9445163.
42. A. Bilal, G. Sun, Y. Li, S. Mazhar and A. Q. Khan, "Diabetic Retinopathy Detection and Classification Using Mixed Models for a Disease Grading Database," in IEEE Access, vol. 9, pp. 23544-23553, 2021, doi: 10.1109/ACCESS.2021.3056186.

43. 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," in IEEE Access, vol. 8, pp. 104292-104302, 2020, doi: 10.1109/ACCESS.2020.2993937.
44. J. Wang, L. Yang, Z. Huo, W. He, and J. Luo, "Multi-Label Classification of Fundus Images With EfficientNet," in IEEE Access, vol. 8, pp. 212499-212508, 2020, doi: 10.1109/ACCESS.2020.3040275.
45. N. Barhate, S. Bhave, R. Bhise, R. G. Sutar, and D. C. Karia, "Reducing Overfitting in Diabetic Retinopathy Detection using Transfer Learning," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), pp. 298-301, 2020, doi: 10.1109/ICCCA49541.2020.9250772.
46. M. Kolla and V. T, "Efficient Classification of Diabetic Retinopathy using Binary CNN," 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), pp. 244-247, 2021, doi: 10.1109/ICCIKE51210.2021.9410719.
47. Y. S. Boral and S. S. Thorat, "Classification of Diabetic Retinopathy based on Hybrid Neural Network," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1354-1358, 2021, doi: 10.1109/ICCMC51019.2021.9418224.
48. R. S. Rajkumar, T. Jagathishkumar, D. Ragul and A. G. Selvarani, "Transfer Learning Approach for Diabetic Retinopathy Detection using Residual Network," 2021 6th International Conference on Inventive Computation Technologies (ICICT), pp. 1189-1193, 2021, doi: 10.1109/ICICT50816.2021.9358468.
49. I. Usman and K. A. Almejalli, "Intelligent Automated Detection of Microaneurysms in Fundus Images Using Feature-Set Tuning," in IEEE Access, vol. 8, pp. 65187-65196, 2020, doi: 10.1109/ACCESS.2020.2985543.
50. V. Lakshminarayanan, H. Kheradfallah, A. Sarkar and J. Jothi Balaji, "Automated Detection and Diagnosis of Diabetic Retinopathy: A Comprehensive Survey", Journal of Imaging, vol. 7, no. 9, p. 165, 2021.
51. W. Alyoubi, W. Shalash and M. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review", Informatics in Medicine Unlocked, vol. 20, p. 100377, 2020.
52. G. Kalyani, B. Janakiramaiah, A. Karuna and L. Prasad, “Diabetic retinopathy detection and classification using capsule networks,” Complex Intell. Syst., vol. 9, no. 3, pp. 2651–2664, 2023.
53. Q. Li, C. Peng, Y. Ma, S. Du, B. Guo, and Y. Li, "Pixel-level Diabetic Retinopathy Lesion Detection Using Multi-scale Convolutional Neural Network," 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), pp. 438-440, 2021, doi: 10.1109/LifeTech52111.2021.9391891.
54. M. S. H. Talukder and A. K. Sarkar, “Nutrients deficiency diagnosis of rice crop by weighted average ensemble learning,” Smart Agricultural Technology, vol. 4, no. 100155, p. 100155, 2023.