Journal of Information Systems Engineering and Management
2025, 10(26s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Rectal Fundus Image Recognition Using Deep Learning Ensemble CNN Models

Dr. T. Nandhini[1], Dr. M. Rajesh Babu[2], Dr. S. Rajalakshmi[3], Mr. M. Rajasekar[4], Dr. S. Uma[5], B.S. Ranjith[6]

[1]Associate Professor, Department of MCA,
Moodlakatte Institute of Technology,
Kundapura, Karnataka, India
Email: nandhini.t13@gmail.com
[2]Professor & Associate Dean, School of Computing,
Rathinam Technical Campus
Coimbatore, India
Email: drmrajeshbabu@gmail.com
[3]Professor, Department of AI & DS,
Dr.NGP Institute of Technology,
Coimbatore, India
Email:mrajislm@gmail.com
[4]Associate Professor, Department of CSE,
VSB College of Engineering Technical Campus, Coimbatore, India
Email: rajasekarcsevsbcetc@gmail.com
[5]Professor, PG In charge, Department of CSE,
Hindusthan College of Engineering and Technology,
Coimbatore, India
Email: druma.cse@hicet.ac.in
[6]PG Student, Department of CSE,
Hindusthan College of Engineering and Technology,
Coimbatore, India
Email: ranjithbs61@gmail.com
Corresponding author email id: nandhini.t13@gmail.com

ARTICLE INFO

Received: 28 Dec 2024

Revised: 21 Feb 2025

Accepted: 01 Mar 2025

ABSTRACT

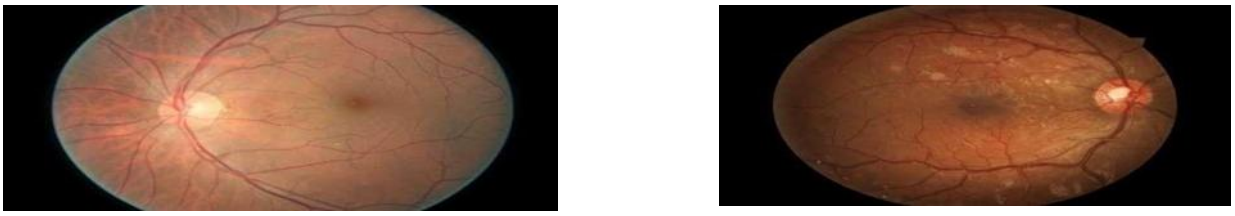
As a vital tool for ophthalmologists to detect potential blinding problems, retinal fundus image analysis is an essential part of medical image analysis. This work looks at retinal fundus images for the purpose of detecting diabetic retinopathy. After every model has been trained independently, the probabilities for every class are totalled to get the ultimate value. The class with the greatest value is called the decision class. Because of this, the proposed methodology consists of the following steps: gathering retinal fundus images; pre-processing images (resizing, contrast enhancement, shade correction, normalization, and image augmentation); feature extraction and classification using a group of specially trained convolutional neural networks; and assessment and evolution of the ideal weights for a more precise diagnosis of diabetic retinopathy in fundus images.

Keywords: Ensemble, Pre-trained Models, Convolutional Neural Networks, Deep Learning, Retinal Fundus Images, Image Classification, Diabetic Retinopathy.

INTRODUCTION

The goal of computer vision is to capture information from an object that resembles what the human eye sees. One of the important fields that greatly benefits from computer vision techniques is biology and medicine. Ophthalmologists examine different regions of the eye to assess a patient's eye health, making it one of the vital disciplines in both medicine and biology. The retina, or the screen where the image is captured, is one of the most important components of the eye and is essential for maintaining excellent eyesight. Retinal damage can have serious consequences, including blindness. Glaucoma, macular degeneration, and diabetic retinopathy (DR) are a few of the most common retinal illnesses that impair vision.

A retinal condition caused by uncontrolled and chronic diabetes is called diabetic retinopathy. Biomarkers like cotton-wool patches, hemorrhages, microaneurysms, and exudates are what define it [11]. To determine the presence or absence of these biomarkers, the retina is scanned using a variety of methods, including fundus photography and optical coherence tomography [8]. Fig. 1 displays an example of retinal fundus pictures obtained using fundus photography.



(a) (b)

Fig. 1: Samples of retinal fundus image pertaining to (a) Normal eye (b) Diabetic Retinopathy affected eye

Diabetic Retinopathy (DR), if not treated in its early stages, may lead to vision loss. Early identification of the disease may help the practitioners provide the right treatment at the right time, thereby aiding in saving the patient ‘s vision. As DR identification is time-consuming and requires expertise, aid from computational intelligence methodologies will be highly appreciated. Identification of the presence or absence of diabetic retinopathy in a retinal fundus image is the focus of this work.

In this aim, numerous computer algorithms [9] have been presented. First, image processing techniques are applied to the retinal fundus images. Following appropriate processing, data mining techniques are applied to the picture characteristics to extract rules that identify the existence or absence of biomarkers, disease, and disease severity. Deep learning techniques are among the various data mining approaches that have gained popularity recently [10]. These techniques have been shown to outperform traditional methods in the majority of computer vision-related applications. Convolutional neural networks, which include convolutions to extract features, pooling to reduce dimensionality, and fully connected layers for classification, are typically used in deep learning techniques for computer vision tasks [11].

This work attempts to examine the provided fundus image and determine whether or not it is impacted by DR. Three deep learning models are offered as an ensemble in this context. The remaining sections of the document are organized as follows: In Section 2, the published works that have already been done on DR identification are reviewed; in Section 3, the suggested methodology for DR identification is explained; in Section 4, the findings showing the effectiveness of the suggested work are discussed; and in Section 5, the work is concluded. It consist of convolutional neural networks, which use pooling to lower dimensionality, convolutions to extract information, and fully connected layers for classification [11]. During training, all of these types of layers' associated weights are optimized. Consequently, the computer vision-focused deep learning models

LITERATURE SURVEY

The works pertaining to the identification of diabetic retinopathy (DR) can be broadly categorized into those that involve biomarker segmentation and those that do not involve biomarker segmentation. Here, the recent works based on deep learning methods that do not involve biomarker segmentation and that have been tested on the EyePACS and MESSIDOR-1 datasets are projected.

A methodology rooted in VGG architecture has been advocated for the identification of DR [13]. The methodology involves a quality assessment module wherein ungradable images are eliminated from further processing and diverted for an expert’s opinion while gradable images are provided to a modified VGG-19 network trained for the identification of DR. Further, a hierarchical classification procedure rooted in a hybrid of Inception-v3 and SVM has been proposed [13]. In this regard, firstly, binary classification is performed to categories images with and without DR. Then, the images detected with DR are further presented to the second stage of classification for categorizing them into various grades. Inception-v3 is employed for the extraction of discriminative features, and the elicited features are provided to the SVM for classification.

Further, a convolutional neural network-based approach has been advocated for classifying DR images [15]. In this methodology, the images are initially pre-processed and passed onto a convolutional neural network for optimizing the weights in the context of DR identification. Then, test images are presented to the optimized model to detect the presence or absence of DR. Another deep learning approach that classifies retinal fundus images into DR and no-DR has been suggested [16]. According to this approach, the images are pre-processed and provided as input to the VGG- D model, which is then fine-tuned to classify DR and no-DR images. Further, a special class coding method is incorporated (useful for multi-class classification) to realize the difference between the predicted and target score during network training.

A methodology rooted in a deep convolutional neural network pre-trained with ImageNet data has been advocated for the binary classification of DR [17]. The methodology involves augmentation of a classification layer to the pre-trained Inception-v3 model. Another work has compared the performance of feature-based machine learning techniques and convolutional neural networks on diabetic retinopathy identification [18]. The outcomes have proven the effectiveness of convolutional neural networks over feature-based machine learning techniques in this regard. Further, on comparing Densenet-121 and Inception-v3 pre-trained models, the latter one provides better DR classification. Further, another method that is rooted in ensembles of deep convolutional networks has been advocated [19]. Here, Inception-v3 and Inception-Resnet-v4 have been employed for training various subsets of the EyePACS dataset. Another ensemble of five deep neural networks has been suggested for this task [20]. In this methodology, the images are initially pre-processed, which involves cropping, adaptive histogram equalization, and local average subtraction. Then, the pre-processed images are used to train five models, namely NTS-Net, SBS Layer, ResNet-50, Densenet- 201, and NASNet, and the predictions from these models are ensembled to derive the final prediction.

Recently, a methodology that exploits a deluge amount of unlabeled data has been advocated [21]. In this methodology, first the unlabeled images are presented to a self-supervised or semi-supervised learning model in the view to formulate a pre-trained model. Then, the pre-trained model is fine-tuned with supervised learning to detect DR and non-DR images. Further, a deep learning model incorporating long-range connections has been suggested for DR detection owing to the spread of DR biomarkers across the fundus area in plaques [22]. In this approach, Inception- v3 is implemented as the backbone, and a long-range unit is augmented to elicit long-range connections in spatial dimensions, thereby aiming at an expressive feature representation that encompasses features that depict local and long-range dependencies.

From the recent studies, it is realized that deep learning methodologies yield improved performance than that yielded by traditional data mining learning models. Further, convolutional neural networks that have been pre-trained on large datasets, when fine-tuned for the current problem in hand, provide enhanced accuracy than convolutional neural networks built from scratch. Furthermore, an ensemble of models provides improved accuracy in the majority of applications. In this work, identification of diabetic retinopathy (DR) is carried out through an ensemble of fine- tuned, pre-trained deep convolutional neural networks. Further, no image is dropped on the basis of gradeability. The entire dataset is considered as such for training, validation, and testing. The following section highlights the materials and the proposed methodology towards classification of DR in fundus images.

MATERIALS AND METHODS

3.0 Dataset

In this work, two publicly available datasets, namely EyePACS [19] and MESSIDOR-1 [24] [25], are utilised for developing and evaluating the proposed models. EyePACS is a retinal image dataset available publicly in the Kaggle repository. The dataset comprises 35126 training images, 10906 public test images (used as a validation set), and 42670 private test images (used as a test set), each labelled with one of the five DR grades, namely 0 to 4. In this work, DR identification is formulated as a binary classification problem, and hence, images marked as 0 are considered to be images without DR biomarkers, while images marked with grades 1-4 are regarded as being affected by DR. Further, another dataset, MESSIDOR-1, is used for assessing the effectiveness of the proposed approach. The dataset comprises 1200 images classified into four categories: normal (grade 0) and DR grades 1 to 3. Similar to that done in the EyePACS dataset, the images are relabelled as either showing DR biomarkers (grades 1-3) or not showing DR biomarkers (grade 0). The proposed methodology for the identification of DR in retinal fundus images is detailed in the following sub-section.

3.1 Proposed methodology for identification of diabetic retinopathy in retinal fundus images

The proposed methodology aims to build an ensemble of supervised deep learning classification models to classify the retinal fundus images as being affected by Diabetic Retinopathy (DR) or not. In this context, the proposed methodology comprises an image collection phase, an image pre-processing phase, a feature extraction and classification phase, and an evaluation phase. The proposed methodology for DR identification from retinal fundus images is depicted in Fig. 2.

Firstly, the training retinal fundus images and the corresponding ground truth (affected by DR or not affected by DR) are provided as input to the image pre-processing phase. During pre-processing, the fundus images are resized, contrast enhanced, shade corrected, and normalized in order to enhance the feature extraction potential and thereby the classification performance. Furthermore, steps related to image augmentation are performed to increase the robustness of the training model. Subsequently, the pre-processed retinal fundus images are presented to the feature extraction and classification phase. Here, three convolutional neural networks that have been pre-trained on a large volume of images from the ImageNet database [26] are customized and fine-tuned to elicit descriptors for DR identification. In this regard, for each of the three convolutional neural networks, deep neural networks are augmented to classify the fundus images as either being affected by DR or not based on the descriptors revealed by its corresponding convolutional neural network. Then, the predictions are aggregated to provide an ensemble prediction. Further, during the evaluation phase, the performance of the model is assessed based on its performance on images that are considered for validation. The various processes involved in the proposed methodology are explained in detail in the following sub-sections.

3.1.1 Image pre-processing phase

The image pre-processing phase involves resizing, contrast enhancement, shade correction and image augmentation. Firstly, the images are resized to a constant size in order to facilitate constant number of inputs for the convolutional neural networks. Images within EyePACS dataset and images between MESSIDOR and EyePACS dataset vary in their resolution. Hence, all these images are converted to a resolution of 512 * 512 pixels. Lower the resolution of the image, lower is the number of parameters for training and thus lower is the complexity of the network and training time. However, lower resolution images may lose important information and thereby impact classification performance negatively. This size (512 * 512) is chosen as a trade-off between the training time and the classification performance. Further, the resized images are contrast enhanced through Contrast Limited Adaptive Histogram Equalization (CLAHE) [27]. In this regard, the three channels of the RGB image are enhanced by improving their contrast. CLAHE algorithm equalizes an image and thereby enhances the contrast. It operates on small sub-images called tiles instead of the entire image. Having applied on all the tiles of the image, the resultants are integrated through bilinear interpolation in the view to discard the false boundaries. The CLAHE algorithm is applied on all the three channels namely, R, G and B. The three contrast enhanced channels are then combined to achieve the contrast enhanced RGB retinal fundus image.

Subsequently, the contrast-enhanced image is subjected to shade correction. This process is performed to make the illumination of the image uniform. For this, the background image is derived by average-filtering the original image with a large-size filter. Then, the background image is subtracted from the original image [28]. Further, the shade-corrected image is normalized by dividing each pixel value by 255, thus allowing a pixel value range of 0 to 1.

After performing the afore-mentioned pre-processing procedures, image augmentation is done [27]. During this process, the original image is transformed by subjecting it to a variety of transformations, thereby yielding transformed versions of the original image. This process aims at introducing more variations in the training set of images, which in turn facilitates a better training process and thereby results in a robust trained model. Incorporating a small amount of transformation to the original image and presenting it for training helps the model gain insight on various perspectives of retinal fundus images belonging to the same class.

For this purpose, Keras ImageDataGenerator, a Python package, is employed [30][31]. It helps in introducing variation in the training images, thereby facilitating the model's ability to generalize better on new, unseen retinal fundus images. The ImageDataGenerator class supports different transformations, namely rotation, flipping, brightening, shifting, zooming, and so on. It works on batches of images. In this work, 16 images are considered at a time in a batch. A few images are selected at random and are subjected to various transformations supported by the ImageDataGenerator class, such as image rotation, brightening, shifting, zooming, and flipping.

In this work, the rotation range is set to 10 as the retinal fundus images could be tilted slightly during capture and the model should get used to such scenarios as well. Further, image brightening is applied. The fundus images may characterize various brightness levels. In order to expose the training model to images of various brightness levels, a range of 0.6 to 1.4 is set. A value below 1 indicates darkening, and a value above 1 denotes brightening the original image. Then, image shifting is applied. The fundus images are shifted horizontally and vertically by 10%. That is, the image is moved top, bottom, and left or right by 10%. Further, flipping is performed on randomly selected images. Both vertical and horizontal flipping are done randomly on selected images. Furthermore, zooming is performed on randomly chosen images within a batch. In this work, the zoom range is set to 0.1. That is, the selected images areeither zoomed in or zoomed out by a maximum of 10%. Thus, the images are pre-processed. A sample of an original image and the outcome after pre-processing are illustrated in Fig. 3.

(a) (b)

Fig. 3: Sample of (a) original image (b) pre-processed image

After pre-processing, the pre-processed images are presented to the feature extraction and classification phase, the details of which are presented in the subsequent sub-section.

3.1.2 Feature extraction and classification phase

The pre-processed images are provided as input to the feature extraction and classification phase. Firstly, the images are passed onto the feature extraction phase, wherein convolutional neural networks are employed to elicit descriptors of retinal fundus images in the context of DR identification. These fundus descriptors are provided as input into the classification phase, wherein a deep neural network is incorporated to classify the retinal fundus image as either affected by DR or not. However, the convolutional neural networks in the feature extraction phase are combined with the deep neural networks in the classification phase and are trained together; therefore, these phases are explained together in this sub-section.

In this work, an ensemble of three convolutional neural networks is proposed to categorize fundus images as either being affected by DR or not. The ensemble is composed of convolutional neural networks [11], namely Inception-v3 [32], VGG-16 [29], and MobileNet-v2 [34], that are pre-trained on millions of images from the ImageNet database [26]. In this work, these convolutional neural networks are augmented with deep neural network layers for DR classification. To each of these three considered convolutional neural networks, two fully connected layers, two dropout layers, and a classification layer are added. All layers of each newly formed network are trained with the pre- processed training images of the dataset. The predictions from these three newly formed networks are aggregated to form an ensemble prediction. Having provided a high-level picture of the feature extraction and classification phase, the details pertaining to each of the considered convolutional neural networks are presented subsequently.

Convolutional neural networks refer to a category of neural networks that specialize in eliciting spatial particulars and salient features from images [11]. It generally consists of convolutional layers, pooling layers, and fully connected layers. It operates on images and performs operations with the view to extract key information from them and provides a feature vector based on which the image can be categorized into one of the many classes in context.

Pre-trained convolutional neural networks are those that have been trained with millions of images in high-end systems for several days, with their weights being optimized for a particular task [35]. These pre-trained models can be reused in some other context to attain better performance than that achieved by a convolutional neural network model developed from scratch. However, the pre-trained networks require fine-tuning in order to adapt to the current context. This work uses three such pre-trained deep convolutional neural network models that have been

trained on the ImageNet database. These networks are tailored to transform them into DR classification networks. The particulars pertaining to three considered convolutional neural networks are presented subsequently.

The first considered pre-trained network is the Inception-v3 model [32], the architecture of which is shown in Fig. 4 [36].

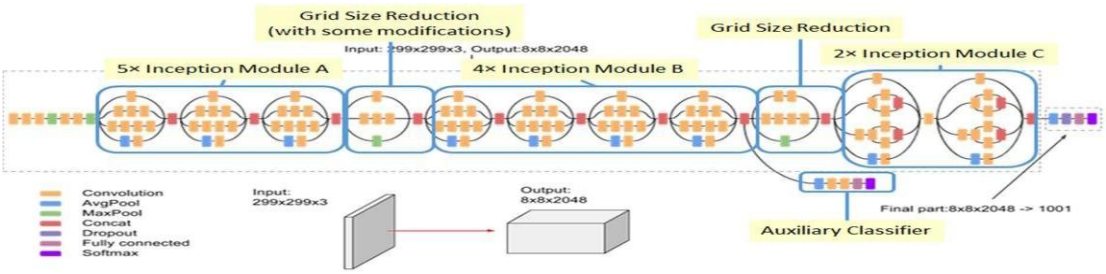


Fig. 4: Architecture of Inception-v3 model

The inception V3 model's general structure is shown in the table 1. Each module's output size serves as the next module's input size.

Table: 1 Inception V3 model's general structure

S. No	TYPE	PATCH / STRIDE SIZE	INPUT SIZE
1	Conv	3×3/2	299×299×3
2	Conv	3×3/1	149×149×32
3	Conv padded	3×3/1	147×147×32
4	Pool	3×3/2	147×147×64
5	Conv	3×3/1	73×73×64
6	Conv	3×3/2	71×71×80
7	Conv	3×3/1	35×35×192
8	3 × Inception	Module 1	35×35×288
9	5 × Inception	Module 2	17×17×768
10	2 × Inception	Module 3	8×8×1280
11	Pool	8 × 8	8 × 8 × 2048
12	Linear	Logits	1 × 1 × 2048
13	Softmax	Classifier	1 × 1 × 1000

It is an advancement of Inception-v1, also known as Google Net, as it has been developed by a team of researchers at Google. In comparison with Inception-v1, factorization into smaller convolutions, spatial factorization into asymmetric convolutions, utilization of auxiliary classifiers, and effective grid size reduction have been incorporated into Inception-v3 in the view of better optimize the network. It characterizes 42 layers. The Inception-v3 model has been trained with the ImageNet database. In this work, the network is augmented with a few more fully connected layers in the view to classify retinal images as either being affected by DR or not.

In the aspect of training the newly formed model for DR classification, the original Inception-v3 architecture along with the weights obtained after several hours of training with the ImageNet database are included. With that, a fully connected layer of 1024 neurons, a dropout layer incorporating 30% dropout, a fully connected layer of 1024 neurons, again a dropout layer marking 30% dropout, and a classification layer encompassing 2 neurons implementing the SoftMax activation function are augmented. These newly added layers hold random weights at

their links. This newly formed network embodying Inception-v3 architecture with the trained weights and the newly added layers with random weights are trained completely with the retinal fundus training images of the

dataset. In this work, the Adam optimizer is utilized for optimizing the weights during training. Further, another convolutional neural network model that is incorporated as a part of the proposed ensemble is the VGG-16 model [33], the architecture of which is depicted in Fig. 5 [37].

Fig. 5: Architecture of VGG-16 model

It has been proposed by researchers from the Visual Geometry Group. The network encompasses 16 layers. It has been optimized for the classification of 1000 objects in the ImageNet database [26]. In this work, the VGG-16 network is considered without the last layer that pertains to the classification of 1000 different objects in the ImageNet database. Instead, two fully connected layers of 1024 neurons with a dropout layer characterizing 30% dropout after each added layer and a final classification layer containing 2 neurons have been included. The VGG-16 architecture with trained weights and the newly added layers with randomly assigned weights are trained further together with the retinal fundus training images.

Furthermore, another model that is included in formulating the ensembles is Mobilenet-v2 [30], the architecture of which is portrayed in Fig. 6 [38].

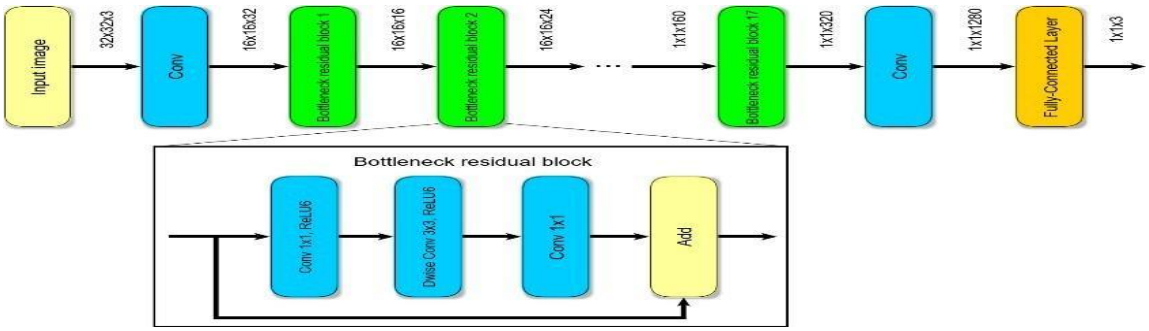


Fig. 6: Architecture of Mobilenet-v2 model

Mobilenet-V2 has been developed by researchers at Google on the concept of inverted residual structures, wherein the residual connections are placed in between the bottleneck layers. It is a light-weight convolutional neural network that has 53 layers. It has also been trained with the millions of images in the ImageNet database in order to classify 1000 categories of images. Here, the trained Mobilenet-v2 model is considered a backbone, and it has been tailored and adapted for identifying the presence or absence of DR biomarkers. In the original Mobilenet-v2 network, the last layer is excluded, and two fully connected layers of size 1024 with 30% dropout after each layer and a final classification layer of size 2 are integrated. The Mobilenet-v2 backbone with trained weights and the newly added layers with random weights are fine-tuned for DR classification, and the weights of the entire network are optimised using Adam optimiser for the task at hand.

Thus, after implementing the afore-mentioned three convolutional neural networks in the context of DR identification in retinal fundus images, they are trained and their weights are optimised. Each of the networks reveals two probabilities corresponding to two possibilities (affected by DR and not affected by DR) for every image. Thus, in the ensemble, there will be three probabilities for each possibility (either affected by DR or not affected by DR). These probabilities are summed. The class that characterises the highest summed value is finalised as the decision class. The models are assessed in the evaluation phase, as detailed in the subsequent sub-section.

3.1.3 Performance evaluation phase

Having implemented the three customized pre-trained convolutional neural networks and developed an ensemble

for DR identification in retinal fundus images, the models and their parameter settings are assessed with the validation and test retinal fundus images provided in the dataset. The performance of a classification algorithm is assessed through true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positive (TP) indicates a DR-affected retinal image being correctly identified as being affected by DR, while true negative (TN) refers to an image not affected by DR being correctly classified as not being affected by DR. On the other hand, FP and FN indicate misclassifications. False Positive (FP) refers to an image not affected by DR being categorized as affected by DR, and False Negative (FN) signifies a condition when an image affected by DR is misclassified as one that is not affected by DR. Based on TP, TN, FP, and FN, many performance measures can be computed, out of which accuracy, sensitivity, and specificity are considered in this work. Further, the Receiver Operating Characteristics (ROC) curve is plotted for the classification outcomes. This curve projects the performance of the learning model at all possible classification thresholds. It is plotted based on the true positive rate (sensitivity) and the false positive rate, computed as the ratio of false positives to the total number of negative instances. The area under the ROC curve (AUC), which measures the two-dimensional area below the plotted ROC curve from (0,0) to (1,1), is computed to quantify the ROC curve.

Thus, in this work, performance is quantified through accuracy, sensitivity, specificity, and area under the ROC curve (AUC). The results associated with various experiments are reported in the following section.

RESULTS AND DISCUSSION

The proposed methodology is evaluated on the EyePACS dataset [23] from Kaggle and the MESSIDOR-1 dataset [24] and [25]. The considered pre-trained models, namely Inception-v3 [32], VGG-16 [29], and Mobilenet-v2 [34], and their ensemble proposed for DR identification in retinal fundus images, are evaluated through evaluation metrics namely accuracy, sensitivity, specificity, and AUC [35]. The methodology is implemented in Python language. The experiments are run on NVIDIA TESLA P100 GPUs provided by Kaggle.

A few parameters are kept constant across all the experiments and trials. Firstly, owing to the high cardinality of images and a medium resolution of 512 * 512, the entire set of training images cannot be processed in a single go. They have to be handled in batches in order to avoid memory issues. In this context, the batch size plays a key role and is set to 16 in this work. Secondly, as mentioned in Section 3, to each of the base pre-trained models, two fully connected layers of size 1024, followed by 30% dropout, and a final classification layer of size 2, are augmented. Further, Kaggle allows a maximum of only 12 hours for a session. Owing to the resolution of the images, each training epoch incurs around 45 minutes or more, and hence the training can be done for only 12 epochs at a time. Then, the final model at the end of each training session (after 12 epochs) is saved, and further training is continued in the next session with the saved model weights of the previous session as the start. Further, the maximum number of epochs is set to 120.

Having set these parameters after careful analysis, each of the three customized pre-trained models is trained independently, and then their predictions are aggregated to derive the final DR class prediction. Firstly, the networks are trained with 35126 training retinal fundus images from the EyePACS dataset [22]. Then, these networks and ensembles are evaluated with the validation and test images of the same dataset. Then, the optimized model is fine- tuned for the MESSIDOR-1 dataset.

In this context, firstly, details pertaining to training each of these three networks and ensembles with the EyePACS dataset are presented, and their performances on validation and test images are reported. Then, the performance is compared against the performance achieved by earlier works considered comparable. Subsequently, the particulars pertaining to fine-tuning the model for the MESSIDOR-1 dataset are narrated. Then, the performance of the fine-tuned model with the test partition of the MESSIDOR-1 dataset is recorded. Further, a performance comparison with earlier works with respect to the MESSIDOR-1 dataset is projected. The following sub-section presents details and outcomes with respect to the EyePACS dataset.

Performance with respect to EyePACS dataset

The three considered models are trained and optimized independently. The Inception-v3 model [32] is trained for 41 epochs. It reaches its maximum validation accuracy at the 36th epoch. In the same manner, VGG-16 [33] is trained for 48 epochs, and the highest validation accuracy is achieved at the 42nd epoch. In the case of Mobilenet-v2 [34], the model achieves the highest accuracy at the 109th epoch and does not show improvement for the next five

epochs. Hence, training is stopped after 114 epochs. The plots of training accuracy versus validation accuracy and training loss versus validation loss over epochs for Inception-v3, VGG-16, and Mobilenet-v2 are shown in Fig. 7 (a) through (f).

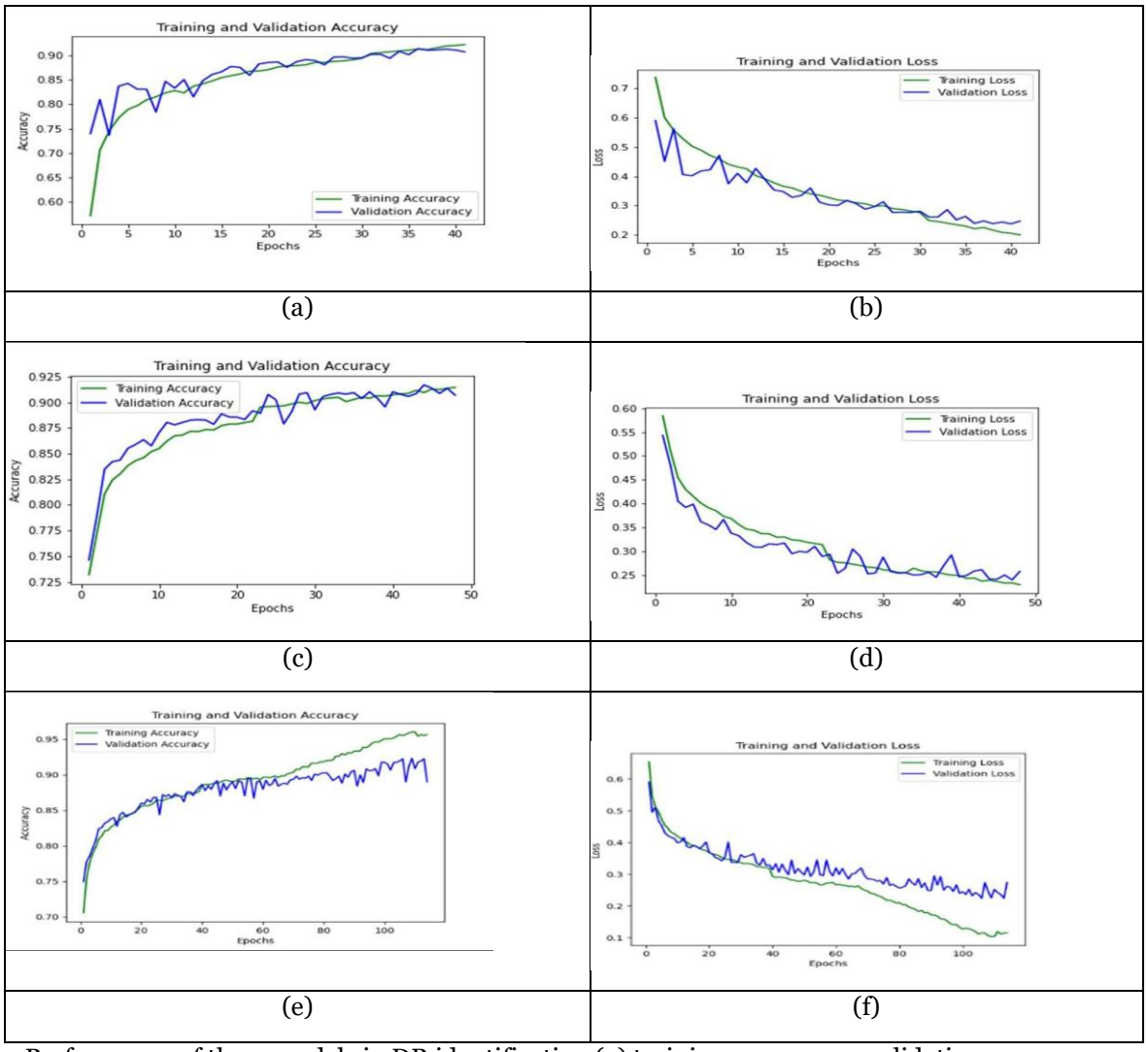


Figure 7: Performance of three models in DR identification (a) training accuracy vs validation accuracy over epochs with respect to Inception-v3 (b) training loss vs validation loss over epochs with respect to Inception-v3 (c) training accuracy vs validation accuracy over epochs with respect to VGG-16 (d) training loss vs validation loss over epochs with respect to VGG-16 (e) training accuracy vs validation accuracy over epochs with respect to Mobilenet-v2 (f) training loss vs validation loss with respect to Inception-v3

Having trained all three models individually to the best of their abilities, their outcomes are aggregated to derive the ensemble model prediction. The outcomes of the individual models and the ensemble and the performance metrics computed based on the outcomes are presented subsequently.

The TP, TN, FP, and FN derived from classification outcomes, the computed performance metrics, plotted ROC curves, and the AUC of the three individual models and their ensemble with respect to the 10906 validation images of the EyePACS dataset are presented in Table 2.

Table 2: Performance of individual models and their ensemble on validation images of EyePACS dataset

Model	TP	FP	FN	TN	Accuracy	Sensitivity	Specificity	AuC
Inception-v3	2247	529	422	7708	91.28	80.94	94.81	0.9401
VGG-16	2019	149	757	7981	91.69	72.73	98.17	0.9375
Mobilenet-v2	2195	256	581	7874	92.33	79.07	96.85	0.9500
Ensemble	2206	89	570	8041	93.96	79.47	98.91	0.9678

Table 2 justifies the better performance of the proposed ensemble when compared to that of the individual models in terms of DR classification. Furthermore, the ROC curves pertaining to DR identification on EyePACS validation images are projected in Fig. 8.

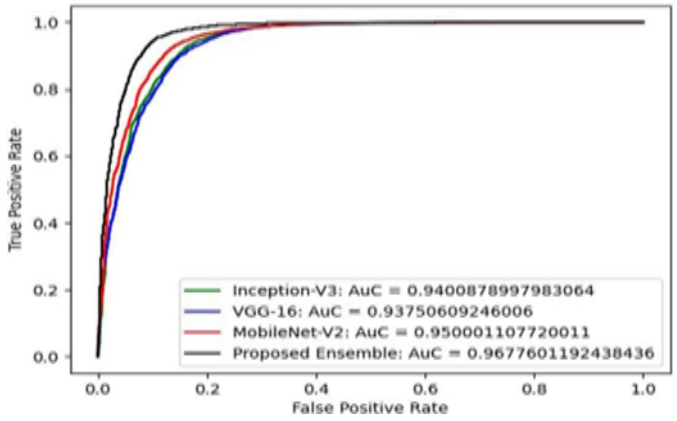


Fig. 8: ROC curves for DR identification in EyePACS validation images

Subsequently, the performance of the individual models and their ensemble is presented for 42670 test images of the EyePACS dataset in Table 3.

Table 3: Performance of individual models and their ensemble on test images of EyePACS dataset

Model	TP	FP	FN	TN	Accuracy	Sensitivity	Specificity	AuC
Inception-v3	8553	2044	2714	29359	88.85	75.91	93.49	0.9133
VGG-16	7818	764	3449	30639	90.13	69.39	97.57	0.9194
Mobilenet-v2	8176	1394	3091	30009	89.49	72.57	95.56	0.9074
Ensemble	8297	559	2970	30844	91.73	73.64	98.22	0.9416

Table 3 demonstrates the enhanced performance of the proposed ensemble when compared to that of the individual models. Subsequently, Fig. 9 portrays the ROC curves pertaining to DR identification in EyePACS test images.

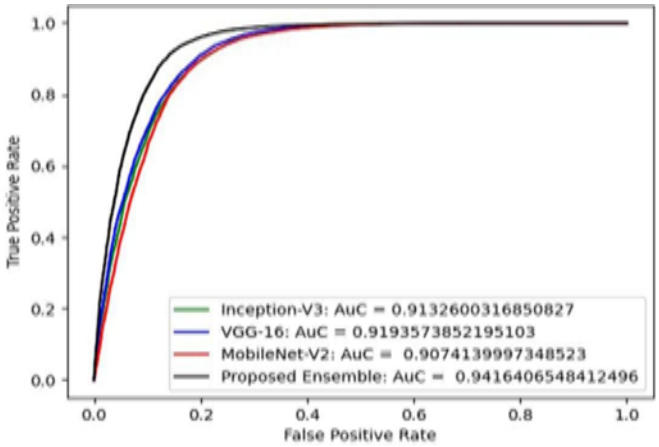


Fig. 9. ROC curve associated with DR identification on EyePACS test set

Furthermore, the performance of the proposed methodology is compared with that of the earlier works on this dataset. It is to be noted that the cardinality of images chosen for training, testing, and validation differs, and hence direct comparison is not possible.

COMPARISON AGAINST STATE OF THE ART METHODS

In this section, we compared the experimental findings with the state-of-the-art methods. As can be seen in the performance comparison of the proposed methodology with that of the earlier works on the EyePACS dataset and

performance comparison with respect to DR classification on the MESSIDOR-1 dataset, our proposed methods outperformed better than the previous works of [11,12,13,14,15,17,18] in terms of accuracy, sensitivity, specificity, and AuC. None of the previous studies did not consider the ensemble of these three CNN models Inception v3, VGG- 16, and MobileNet-v2. Having presented the recent works that have been evaluated with the EyePACS dataset and/or MESSIDOR-1 dataset, a few very recent works on diabetic retinopathy detection, though do not pertain to EyePACS and MESSIDOR-1 datasets, are projected here. Our solution demonstrated the best performance in the experiments, indicating that our model may work well for DR detection in the real world. Additionally, it would be used to address other relevant issues in the field of medical image processing.

A wrapped approach using Unlabelled data for diabetic retinopathy diagnosis using methodology Semi-supervised learning followed by supervised learning with backbone as ResNet. This work applied the EyePACS dataset, it was produced an accuracy is **86.40**, and compare to our proposed model novelty ensemble of the Inception- v3, VGG-16, and MobileNet-v2 shows high accuracy is **91.73**, Sensitivity is **73.64**, Specificity is **98.22**, and AUC of **0.94916**.

A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence using methodology Inception-v3 integrated with a long-range unit [18] - 2023. This work applied the MESSIDOR-1 dataset and produced an Accuracy is **92.1**, Sensitivity is **93.1**, Specificity is **92.5**, and AUC of **0.967** and compare to our proposed model (novelty ensemble of the Inception-v3, VGG-16, and MobileNet-v2) produced high Accuracy is **97.77**, Sensitivity is **97.45**, Specificity is **98.17** and AUC **0.9957.**

CONCLUSION

Utilization of computational techniques for automatic analysis of retinal fundus images with a view to diagnosing abnormalities has been a profound area of research in recent times. This work focuses on identifying if the presented retinal fundus image is affected by Diabetic Retinopathy (DR) or not. In this regard, the proposed methodology involves retinal fundus image collection, retinal fundus image pre-processing encompassing resizing, CLAHE contrast enhancement, shade correction, normalization, and augmentation, feature extraction, and a classification phase implemented through an ensemble of three customized pre-trained models, namely Inception-v3, VGG-16, and Mobilenet-v2, performance evaluation, and the evolution of the best classification weights for DR identification. The proposed methodology achieves an accuracy and AUC of 93.96% and 0.9678 on the EyePACS public test set (validation images in this work) and 91.73% and 0.9416 on the EyePACS private test set images (test set in this work). Further, the model has been fine-tuned to identify Diabetic Retinopathy in the MESSIDOR-1 dataset. In this context, the proposed methodology achieves an accuracy and AUC of 97.77% and 0.9893 on the test partition of the MESSIDOR-1 dataset, respectively. Thus, the proposed methodology exhibits justifiable performance, making it a good candidate for integration in eye hospitals to aid ophthalmologists in DR screening.

The major derivations from the work are: contrast enhancement and shade correction expose the DR biomarkers more prominently and thereby aid in enhancing the recognition potential of the learning model; image augmentation makes the training process more efficient thereby facilitating better recognition capabilities; pre-trained deep learning models already trained with some kind of images get better tuned to DR identification in retinal fundus images; and an ensemble of deep learning models achieves better prediction performance when compared to that of the individual models.

This work can be further extended in the future by making this a multi-class problem, thereby identifying severity grades of Diabetic Retinopathy. Further, attention mechanisms can be incorporated to improve the identification potential of the learning model. Furthermore, the developed model can also be adapted and fine-tuned for identifying other retinal diseases such as glaucoma, macular degeneration, and so on.

References

[1] Mythili, S., Rajesh Babu, M., Naveen Sundar, G., Uma, S., Thangavel, P., and Anitha, S., “Swaram Extraction from Veena Tunes using Deep Learning Algorithms”, Journal of Information Systems Engineering and Management, Vol.10, No. 21s, 2025.

[2] Balaji Shanmugam, V., Rajesh Babu, M., Kalpana, V., Uma, S., Mohana Sundaram, S., and Devika Anil., “An Enhanced RNN-LSTM Model for Image Classification using Deep Learning Techniques”, Journal of Information Systems Engineering and Management, Vol.10, No. 21s, 2025.

[3] Sruthi S Madhavan., Kala, I., Rajesh Babu, M., Uma, S., Brindha, D., and Karthika, S., "Integration and Innovation Path Analysis of Enterprise Marketing Data Management Based on Deep Learning", Journal of Information Systems Engineering and Management, Vol.10, No. 21s, 2025.

[4] R. Venkatesan, Abdul Karim, Rajesh Babu, M., S. Uma, S. Rajeswari and Akash Kumar, "Rectal Fundus Image Recognition Using Deep Learning Ensemble CNN Models", Journal of Information Systems Engineering and Management, Vol.10, No. 21s, 2025.

[5] B. Anuradha, M. Kasi Pandian, Rajesh Babu, M., S. Uma, S. Palanisamy and B. Venkatraj, "Web Application Development for Product Label Reading using Machine Learning Techniques," Journal of Information Systems Engineering and Management, Vol.10, No. 21s, 2025.

[6] K. M. Kim, T. Y. Heo, A. Kim, J. Kim, K.J. Han, J. Yun, J.K. Min, "Development of a fundus image-based deep learning diagnostic tool for various retinal diseases", Journal of Personalized Medicine. Vol. 11, No. 5, pp. 321. https://doi.org/10.3390/jpm11050321.

[7] R. G. Ramani, B. Lakshmi, "Automatic diabetic retinopathy detection through ensemble classification techniques", 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 14 December 2017, pp. 1-4. 10.1109/ICCIC.2017.8524342

[8] M. D. Abramoff, M. K. Garvin, M. Sonka, "Retinal imaging and image analysis", *IEEE reviews in biomedical engineering*. Vol. 3, pp. 169-208, 2010. 10.1109/RBME.2010.2084567

[9] N. Patton, T.M. Aslam, T. MacGillivray, U. Deary, B. Dhillon, R.H, Eikelboom, K. Yogesan, I.J. Constable, "Retinal image analysis: concepts, applications and potential", *Progress in Retinal and Eye Research*, Vol. 25, No. 1, 2006, pp. 99-127. https://doi.org/10.1016/j.preteyeres.2005.07.001

[10] M. Badar, M. Haris, A.Fatima, "Application of deep learning for retinal image analysis: A review", *Computer Science Review*, Vol. 35, 2020, p. 100203. https://doi.org/10.1016/j.cosrev.2019.100203.

[11] K. O'Shea, R. Naah, "An introduction to convolutional neural networks", *arXiv preprint arXiv:1511.08458*, 2015. https://doi.org/10.48550/arXiv.1511.08458

[12] Y. Cao, T. A. Geddes, J.Y. Yang, P. Yang, "Ensemble deep learning in bioinformatics", *Nature Machine Intelligence*, Vol. 2, No. 9, 2020, pp. 500-508. https://doi.org/10.1038/s42256-020-0217-y.

[13] A. Rakhlin, "Diabetic retinopathy detection through integration of deep learning classification framework", *BioRxiv*, 2017, p. 225508. https://doi.org/10.1101/225508.

[14] U. Shrivastava, M.V. Joshi, "Automated multiclass diagnosis of diabetic retinopathy using hierarchical learning", *11th Indian Conference on Computer Vision, Graphics and Image Processing*, 18 December 2018, pp. 1-7. https://doi.org/10.1145/3293353.3293412.

[15] A.R. Choudhury, D. Bhattacharya, A. Debnath, A. Biswas, "An integrated image processing and deep learning approach for diabetic retinopathy classification", *Second International Conference on Advances in Computational Intelligence, Security and Internet of Things, ICCISIoT 2019,* Agartala India, 13-14 December 2019, Proceedings 2 2020, pp. 3-15. https://doi.org/10.1007/978-981-15-3666-3_1

[16] A. Kwasigroch, B. Jarzembinski, M. Grochowski, "Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy", *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 9 May 2018, pp. 111-116. 10.1109/IIPHDW.2018.8388337

[17] M.T. Hagos, S. Kant, "Transfer learning-based detection of diabetic retinopathy from small dataset", *arXiv preprint arXiv:1905.07203*, 2019. https://doi.org/10.48550/arXiv.1905.07203.

[18] J. Goncalves, T. Conceicao, F. Soares, "Inter-observer reliability in computer-aided diagnosis of diabetic retinopathy", *12th International Joint Conference on Biomedical Engineering Systems and Technologies, BOISTEC*, 2019, pp. 481-491. 10.5220/0007580904810491.

[19] G. Saxena, D.K. Verma, A. Paraye, A. Rajan, A. Rawat, "Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public datasets", *Intelligence-based Medicine*, Vol. 3, 2020, p. 100022. https://doi.org/10.1016/j.ibmed.2020.100022. *and Analysis (MIUA 2019)*, Liverpool, UK, 24-26 July 2019, Proceedings 23 2020, pp. 242-253. https://doi.org/10.1007/978-3-030-39343-4_21

[20] X. Zhang, Y. Kim, Y.C. Cung, S. Yoon, S.Y. Rhee, Y.S. Kim, " A wrapped approach using Unlabelled data for diabetic retinopathy diagnosis", *Applied Sciences*, Vol. 13, No. 3, 2023, p.1901. https://doi.org/10.3390/app13031901.

[21] X. Luo, W. Wang, Y. Xu, Z. Lai, X. Jin, B. Zhang, D. Zhang, " A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence", *CAAI Transactions on Intelligence Technology*, 2023. **https://doi.org/10.1049/cit2.12155**

[22] Diabetic Retinopathy Detection – Identify signs of diabetic retinopathy in eye images, 2015. Available Online: https://www.kaggle.com/competitions/diabetic-retinopathy-detection

[23] E. Decenceire, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, "Feedback on a publicly distributed image database: The Messidor database", *Image Analysis and Stereology*, Vol. 33, No. 3, 2014, pp. 231-234. 10.5566/ias.1155

[24] Messidor-ADCIS. Available Online https://www.adcis.net/en/third-party/messidor/

[25] J. Dneg, W. Deng, R. Socher, L.J. Li, K. Li, L. Fei-Fei, "Imagenet: A large-scale hierarchical image database", *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 20 June 2009, pp. 248-255. 10.1109/CVPR.2009.5206848

[26] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, b. ter Haar Romeny, J.B. Zimmerman, K. Zuiderveld, "Adaptive histogram equalization and its variations", *Computer Vision, Graphics and Image Processing*, Vol. 39, No. 3, 1987, pp. 355-368. https://doi.org/10.1016/S0734-.189X(87)80186-X.

[27] C. Pereira, D. Veiga, J. Mahdjoub, Z. Guessoum, L. Goncalves, M. Ferreira, J. Monteiro, "Using a multi-agent system approach for microaneurysm detection in fundus images", *Artificial Intelligence in Medicine*, Vol. 60, No. 3, 2014, pp. 179-188. https://doi.org/10.1016/j.artmed.2013.12.005

[28] C. Shorten, T.M. Khoshgoftaar, "A survey on image data augmentation for deep learning", *Journal of Big Data*, Vol. 6, No. 1, 2019, pp. 1-48. https://doi.org/10.1186/s40537-019-0197-0.

[29] F. Chollet, "Building powerful image classification models using very little data", *Keras Blog*. Vol. 5, pp. 90-95.

[30] A. Gulli, S. Pal, *Deep Learning with Keras*, Packt Publishing Ltd., 2017.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the inception architecture for computer vision", *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826. 10.1109/CVPR.2016.308

[32] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556. 2014. https://doi.org/10.48550/arXiv.1409.1556

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks", *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510-4520. 10.48550/arXiv.1801.04381.

[34] L. Torrey, J. Shavlik, "Transfer Learning", *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques*, IGI Global, 2010, pp. 242-264.

[35] OpenGenus IQ, Inception V3 Model Architecture. Available Online https://iq.opengenus.org/inception-v3-model-architecture/

[36] T. Pakhale. Transfer Learning using VGG16 in Pytorch. Analytics Vidhya. 2021. Available Online https://www.analyticsvidhya.com/blog/2021/06/transfer-learning-using-vgg16-in-pytorch/

[37] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, & E.T. Matson, "Real-time and accurate drone detection in a video with a static background", *Sensors*, Vol. 20, No. 14, pp.3856, 2020. https://doi.org/10.3390/s20143856

[38] D.M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation", *arXiv preprint arXiv:2010.16061*, 2020. https://doi.org/10.48550/arXiv.2010.16061