

Optical coherence tomography angiography image classification and analysis of diabetic retinopathy, using Wasserstein generative adversarial network augmentation

Pranali Pradeep Hatode^{1,2}, Maniroja Edinburg¹

¹Department of Electronics and Telecommunication Engineering, Thadomal Shahni Engineering College, University of Mumbai, Mumbai, India

²Department of Electronics and Telecommunication Engineering, K J Somaiya Institute of Technology, University of Mumbai, Mumbai, India

Article Info

Article history:

Received Mar 25, 2024

Revised Jul 15, 2024

Accepted Aug 6, 2024

Keywords:

Diabetic retinopathy
Non-proliferative diabetic retinopathy
Optical coherence tomography angiography
Proliferative diabetic retinopathy
Wasserstein generative adversarial network

ABSTRACT

Deep learning algorithms effectively work with, a significant amount of data. trained on small datasets tend to have poor generalization. Data augmentation techniques can be used to make better use of existing training data, improving the applicability of deep learning methods. However, traditional data augmentation methods often produce limited additional credible data. The deep learning approach's performance can be enhanced by generating new data by employing generative adversarial networks (GANs). Although GANs have been extensively used to improve the performance of convolutional neural networks (CNNs), there has been relatively less research on data augmentation methods specifically for GAN training. This study focuses on using a Wasserstein GAN (WGAN) architecture for generating synthetic optical coherence tomography angiography (OCTA) images of diabetic retinopathy to aid in the detection of different types of diabetic retinopathy diseases, including proliferative diabetic retinopathy (PDR), Severe non-PDR (NPDR), Moderate NPDR, and Mild NPDR. WGAN, provides the generator with a more informative learning signal, making training more stable, particularly in high-dimensional spaces. The trained WGAN model is saved in .h5 file format (HDF), converted to portable network graphics (PNG) image format, and then classified into different categories of diabetic retinopathy using a ResNet50 model with various fine-tuning methods. The proposed model has demonstrated better results than the previous study. 99.95% accuracy is exhibited.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Pranali Pradeep Hatode

Department of Electronics and Telecommunication Engineering, K J Somaiya Institute of Technology,

University of Mumbai

Mumbai, India

Email: phatode@somaiya.edu

1. INTRODUCTION

Deep learning (DL) has demonstrated promising performance with sufficient training data, but limited-sized datasets continue to cause overfitting, especially in medical imaging. Augmenting data efficiency can improve DL techniques, but this may introduce noise and affect data quality. This study focuses on analyzing and categorizing diabetic retinopathy. The International Diabetes Federation reported 537 million adult diabetics in 2021, projected to rise to 643 million by 2030 and 783 million by 2045. Diabetic retinopathy poses a significant risk of vision loss and life-altering complications, including kidney

and cardiovascular diseases [1]. There are primarily two types of diabetic retinopathy: non-proliferative diabetic retinopathy (NPDR), also known as non-proliferative diabetic retinopathy, which is a relatively milder and symptomless type, and proliferative diabetic retinopathy (PDR), the stage of DR that develops the quickest and shows novel and unusual development of retinal blood vessels. Before advancing to PDR, NPDR passes through mild NPDR, moderate NPDR, and severe NPDR.

This work aims to categorize DR optical coherence tomography angiography (OCTA) images into the above-mentioned types using OCTA images with deep learning methods for classification. However, the unlabeled DR-OCTA images obtained from the IEEE data portal were insufficient in number for effective DL model training. To address this, Wasserstein generative adversarial network (WGAN) was used to enhance the dataset.

WGAN was chosen over the regular augmentation technique of rotating, and flipping, for its ability to train the critic to optimality, eliminating the need for proper balancing of the generator and discriminator's capacities [2]. After training, the model-generated .h5 files were stored in a hierarchical data format (HDF) due to the large volume of data. The trained layers and weights were then loaded into the PNG image generator model to produce augmented PNG images as .png images suit best to deep learning models. These images are employed in the ResNet50, a deep learning technique, for the analysis and classification of DR into different groups. The accuracy and loss values achieved from the ResNet50 model were not the best, so the classification model was fine-tuned by varying various hyperparameters, including freezing and unfreezing of layers, removal of classification head, addition of new classification layers, number of epochs, batch size, learning rate, drop-out layer, and weight-decay. The research utilizes the ResNet50 model to predict various categories of DR for each random unlabeled OCTA image. The framework mentioned above of this research work was decided after a review of various previously done works.

The previous research provides concise summaries of tasks related to improving OCTA images. Yuan *et al.* [3] have created a super-resolution angiogram reconstruction generative adversarial network (SAR-GAN), which is a deep learning technique that enhances the quality of en-face OCTA images. This study used 50 OCTAs from healthy volunteers. Lyu *et al.* [4] proposed automated augmentation for domain generalization (AADG), which utilizes a data manipulation approach to generalize the domain and address the issue of domain gap between training and test datasets. They used OCTA-500 and ROSE (OCTA) for this purpose. Chinkamol *et al.* [5] have proposed OCTAve, a new scribble-based weakly supervised framework. It combines self-supervised deep supervision with adversarial deep supervision using the public datasets OCTA-500 and ROSE. The framework outperforms other scribble-based weakly supervised methods and even the state-of-the-art fully supervised method for the 2D en-face OCTA vessel segmentation task. Hua *et al.* [6] proposed a network called TFA-Net for grading the severity of diabetic retinopathy (DR). This network uses a small number of fundus and widefield SS-OCTA images. It works in two steps: first, it coordinates features from different image types using shared convolution kernels, and second, it utilizes the RCA stream. In their work, Zhou *et al.* [7] propose a new network structure called three-dimensional convolution attention module-ResNet (TCAM-ResNet). The author uses three-dimensional OCT images to classify them into two retinal diseases: AMD and DR. The process involved using around 200 images for the classification. Alavee *et al.* [8] propose a 5-layer convolutional neural network (CNN) with channel size gradually increasing from 8 to 128, on a total of 3,662 fundus images of all five DR categories and have achieved an accuracy of 95.27%. Jabbar *et al.* [9] propose GoogleNet+ResNet in the research work and perform this technique on 35,126 Fundus images that are further augmented to increase 3.6 times, from Kaggle to achieve 94% accuracy. Zekai *et al.* [10] implemented a multi-branching temporal convolutional network with tensor data completion (MB-TCN-TC) to categorize DR into five levels of the disease on around 63 million patient data and achieved 0.949 AUROC as the data used was unbalanced. Jagadesh *et al.* [11] adopt an improved contoured convolutional transformer (IC2T) to segment the fundus image. A detection module for optical disc (OD) and blood vessels (BV) and a dual convolutional transformer block that combines local and global contexts to make reliable associations are developed. Furthermore, a second-stage improved coordination attention mechanism (ICAM) network identifies retinal biomarkers for DR such as microaneurysms, hemorrhages, and exudates. With an average accuracy of 96%, 97%, and 98% on EyePACS-1, Messidor-2, and DIARETDB0 datasets, respectively. Li *et al.* [12] propose a U-Net-based segmentation model to label the boundaries of large retinal vessels and the foveal avascular zone in OCTA images. The experiments were conducted on 301 OCTA images. An accuracy of 93.1% detected the DR and non-DR images. Le *et al.* [13] implemented VGG16 to classify 180 OCTA images into DR and NoDR and achieved an accuracy of 87.27%. Ryu *et al.* [14] have implemented a CNN model for the detection of different categories of DR using 301 OCTA different-size images and achieved 90.8% accuracy for 3×3 mm² and 93.3% accuracy for 6×6 mm² OCTA images. Zang *et al.* [15] in this research work the author has proposed a CNN-based densely and continuously connected neural network with adaptive rate dropout (DcardNet) designed for the DR classification on 303 OCTA images and achieved an accuracy of 94.3%.

A lot of research has been carried out on DR detection using fundus images but it was noted from all the referrals that there was very little work done in recent years, on OCTA images due to their scarcity. Ophthalmologists have pointed out that OCTA can show the very early stages of diabetic retinopathy (DR) as it displays the fine microvascular changes in the retina, which is the earliest stage of diabetic retinopathy, whereas fundus images fail to show these minute early microvascular changes. Also, whatever research was conducted on OCTA images for DR analysis used a small number of images, which posed a challenge for deep learning models. Many authors addressed this by employing general augmentation techniques such as flipping and rotating, but these compromised the data quality and ultimately affected the accuracy of the model. No research was found to provide a correct diagnosis for the unlabeled image, which this research offers. The proposed framework aims to address all these issues. ResNet50 was selected because it exhibits the most optimal results, due to its ability to alleviate the gradient loss problem. Before finalizing ResNet50 for the research, the OCTA images were analyzed using a five-layered custom CNN model without data augmentation, as well as with general, GAN, and WGAN augmentation techniques. The results obtained were unsatisfactory, leading to the implementation of ResNet50 with GAN and WGAN augmentation, with the latter combination producing better results. Further improvement was achieved by fine-tuning the hyperparameters of the models. The result section discusses the outcomes of all the mentioned combinations.

The research work thoroughly discusses the techniques used and the data in the method section. The Results and Discussion section compares the accuracies of the related work with the results achieved from the implemented deep learning techniques. The outcomes of the proposed framework are compared with previous research to justify the selected techniques, which have enabled accurate disease diagnosis with minimal data in a short time span. The limitation of this research is that it relies on data from the IEEE data portal, which is prone to noise. As a result, when applying the same framework to real-time data, the model may need some additional adjustments.

2. METHOD

The research begins with an explanation of the WGAN data augmentation technique, followed by a detailed discussion of the PNG image generator model. Subsequently, the ResNet50 classification model and prediction model are discussed in detail. For this research, 800 OCTA images of five categories were obtained from the IEEE data portal and verified by ophthalmic experts. The study used a WGAN generation model to create 19 instances of images for each category, and a PNG image generation model then produced 500 images for each instance. This resulted in a total of 47,500 images available for classifying and predicting diabetic retinopathy categories.

2.1. WGAN generator and critic architecture

Wasserstein generative adversarial network is known by its acronym, WGAN. It is a particular kind of GAN or generative adversarial network. The goal of WGANs is to get around some of the drawbacks of conventional GANs, especially training stability and mode collapse it achieves this by calculating the difference between two probability distributions using Earth Mover's distance.

The proposed WGAN architecture as shown in Figure 1 [16] uses a batch of 100 real training images and a random noise vector for improved model generalization and performance. A five-layered convolution network is designed, with random noise samples passed through convolutional layers and convolutional 2D layers with filters [17]. The LeakyReLU activation function is employed to overcome dying neurons [18]. The final Conv2D layer configures convolutional layers with Tanh activation for deeper network learning.

The WGAN model as shown in Figure 1 critic architecture consists of two convolutional neural networks with 32 and 64 filters. The model concatenates the generator and critic, with most layers frozen during the training phase to improve the generator's performance based on the critic's input, enhancing sample quality and stabilizing the training process.

2.2 PNG image generator

This work also proposes a PNG image generator. One of the hierarchical data formats (HDF) that WGAN produces is the *.h5* format. It is employed to store enormous volumes of data as multidimensional arrays. PNG is one of the image formats that work best for the mentioned deep learning technique as it is transparent and lossless, it can be used for images that need to be sharp and retain details without exhibiting compression artefacts although it typically has bigger file sizes. Figure 2 shows the flow chart of the PNG image generator. The image instances from the WGAN which are in *.h5* format are uploaded to the model. The dictionary is created to store the generated images. Further, the PNG image generator model is loaded followed by the creation of the array of total samples of noise to train the WGAN model. The loaded model

is called and the path of the image instances from WGAN is loaded to the model. Later the function to generate the images is defined to yield 500 images per category. After the completion of all the epochs, the zipped image folder can be downloaded.

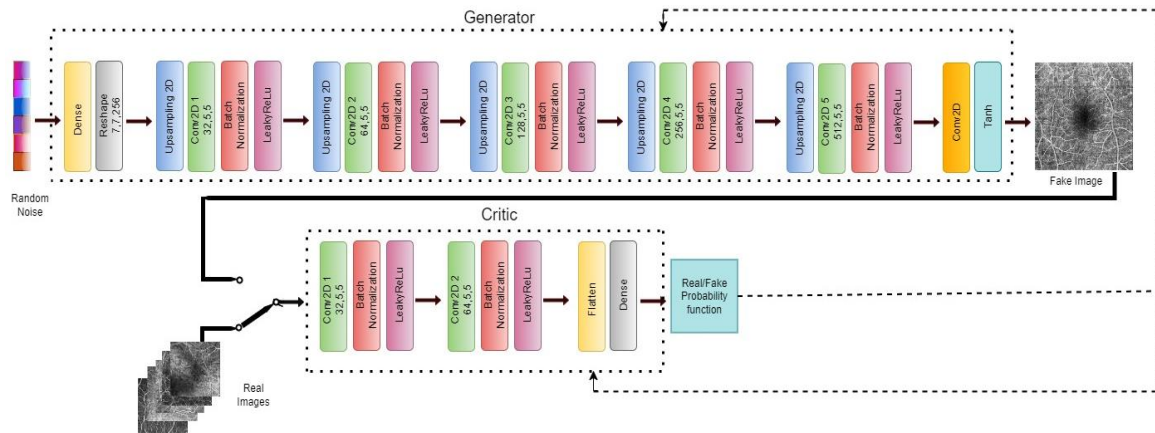


Figure 1. Wasserstein GAN architecture

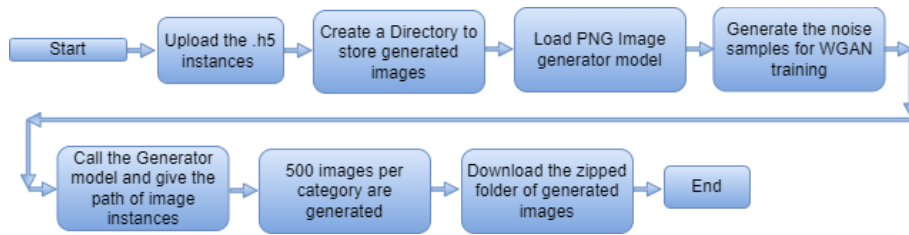


Figure 2. PNG image generator flow chart

2.2.1. PNG image generator output

PNG image generator generates 500 OCTA images from every instance of WGAN. These instances are in HDF formats, which is basically an array of the weights of all the augmented data. PNG image generator converts it to the data in PNG format. Figure 3 shows the sample images of all five categories in PNG format.



Figure 3. PNG image generator output in .png format

2.3. ResNet50 classification workflow

A pre-trained ResNet50 which is the convolutional neural network architecture that belongs to the Resnet family is used in this research work. Figure 4 shows the flow chart of the pre-trained ResNet50 model [19], of which all the layers are fixed but, in this research, the model weights are trained instead of using pre-trained weights that are ImageNet weights. Other than this the final layer is added with a Conv2D layer to classify the images into five categories with SoftMax activation. The OCTA images in PNG format from the

PNG image generator are uploaded to the model. For classification purposes, a batch size of 16 is used. Around 70% which is approximately 328 images are used to train the model, 20% which is around 93 images are used for testing purposes, and around 80 images are used for validation purposes. After that, the pre-trained ResNet50 model is defined and compiled to get the loss function and accuracy at 20 epochs. In the result obtained it was observed that the validation accuracy went under several fluctuations at certain epochs and the loss increased drastically at epochs between 15-17.

So, to improve the results, scheduling the learning rate was incorporated into the model, where the model refers to the practice of adjusting the learning rate during training under specific conditions. This improved the result to some extent. Then secondly the first 150 layers of the model were frozen to avoid catastrophic forgetting of the model and after this the epoch size was increased to 100, yielding good results [20].

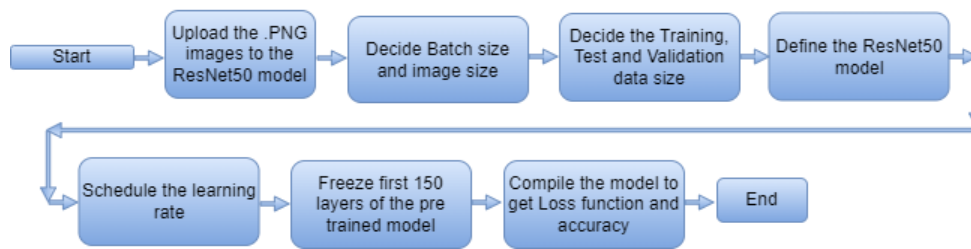


Figure 4. ResNet50 classification model flowchart

2.4. ResNet50 prediction workflow

The workflow of a pre-trained ResNet50 model used for the prediction of the classes or the five categories of the DR is shown in Figure 5. Here the *.h5* file of the classifier model is loaded to the prediction model. At the same time, all categories of OCTA diabetic retinopathy images are uploaded to the model. It is been resized to 224×224, adjusted the shape, and aligned to the model to get its category predicted.

The implementation of this method will provide a convenient diabetic retinopathy classification and prediction system. This will greatly assist ophthalmologists, as the proposed model and framework can quickly and accurately analyze DR-OCTA images. Its capability of detecting the disease at a very early stage will be of great benefit to patients, potentially saving them from serious vision problems or blindness. This system will make it easier for ophthalmologists to start the treatment quickly.

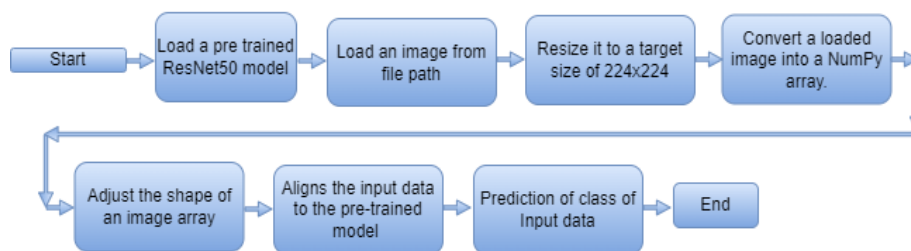


Figure 5. ResNet50 prediction model flowchart

3. RESULTS AND DISCUSSION

The comparison of the findings and the outcomes of the previous research work are illustrated in Table 1, and its analysis shows that although the deep learning techniques are very efficient for the classification, segmentation and prediction of complicated medical images, the appropriate augmentation technique enhances the performance of the model and eventually its accuracy.

Before opting for the WGAN augmentation technique for data enhancement and using ResNet50 for classifying and predicting DR and its five categories, it was analyzed that the performance of the CNN model with different data sizes, augmentation techniques (general augmentation, GAN, and WGAN), and epoch sizes. The results achieved from various combinations are presented in Table 2. It is evident that as the data size increases, data augmentation becomes necessary for better model performance. For instance, with

100 images and no augmentation, 43.44% accuracy was achieved. When the images were augmented to 800 images with simple techniques like flipping and rotating, the accuracy improved to 48.97%.

However, increasing the number of images and adding two more layers to the CNN, while also increasing the number of epochs to 150, resulted in a decrease in accuracy to 20.8%, even with GAN-augmented images. This indicates that the model was underfitting. Subsequently, increasing the number of images to 47,500 only resulted in a marginal improvement in accuracy to 37.57%. Further attempts to improve accuracy involved replacing GAN with WGAN, which led to an accuracy of 40.30%.

Table 1. Related research work findings and outcomes table

Reference. No.	Method	Augmentation	Image type	Number of images	Classification	Accuracy
[4]	AADG	Automated augmentation	OCTA images	445	Segmentation of images in different domains	94.75%
[5]	Self-supervised deep supervision (SSDS)	Self-augmented	OCTA images	229	Segmentation of images	81.42%
[6]	Twofold feature augmentation (TFA)	Augmented	OCTA images	297	DR severity	94.8%
[7]	TCAM-ResNet	-	OCTA images	376	AMD, DR and NORMAL	83.3%
[8]	5-layer CNN	Regular augmentation technique	Fundus images	3,662	DR, Mild-DR, Moderate-DR, Severe-DR and PDR	95.27%
[9]	GoogleNet+ResNet	Regular augmentation technique	Fundus images	35,126	DR, Mild-DR, Moderate-DR, Severe-DR and PDR	94%
[10]	MB-TCN-TC	-	Fundus images	414,199	DR and No-DR	AUROC-0.949 and AUPRC-0.793
[11]	IC2T	-	Fundus, EyePACS-1, Messidor-2, DIARETDB0	8,980, 1,748, 130	DR, Mild-DR, Moderate-DR, Severe-DR and PDR	96%, 97%, 98%
[12]	U-Net+isolated concatenated block (ICB)	-	OCTA images	301	Segmentation	92%
[13]	CNN	-	OCTA images	131	DR and No-DR	97%
[14]	CNN	-	2 sized OCTA images	301	Normal, Mild-DR and PDR	$3 \times 3 \text{ mm}^2$ -90.8% $6 \times 6 \text{ mm}^2$ -93.3%
[15]	Densely and continuously connected neural network with adaptive rate dropout (DcardNet)	-	OCTA images	303	No-Dr and DR, No-Dr, NPDR and PDR, No-DR, Mild-Moderate-NPDR, Severe-NPDR, PDR	98.1%, 87.2%
[16]	Proposed method WGAN+ResNet50	WGAN augmentation	OCTA images	800	DR, Mild-DR, Moderate-DR, Severe-DR and PDR	74%, 99.75%

Table 2. Deep learning techniques, data set, augmentation, epochs, and accuracy comparison table

DL Technique	Data Set (OCTA Images)	Augmentation technique	Epoch	Accuracy
3-layer CNN	100	---	30	43.44%
3-layer CNN	800	Yes	30	48.97%
5-layer CNN	12,500	GAN	150	20.8%
5-layer CNN	47,500	GAN	150	37.57%
5-layer CNN	47,500	WGAN	150	40.30%
ResNet50	47,500	GAN	150	62.48%
Proposed model (fine-tuned)	47,500	WGAN	150	99.95%

It became clear that increasing the number of layers in the deep learning model does not always yield better results, rather can lead to gradient loss, despite implementing effective augmentation techniques. Therefore, we replaced the CNN with ResNet50, operating with the same dataset and epoch size, but using GAN augmentation. This change improved the accuracy to 62.48%. Further, replacing GAN with WGAN resulted in an accuracy of 84.57%. By fine-tuning the hyperparameters of the ResNet50 model, 99.95% accuracy was achieved.

The process of fine-tuning the model starts as described and is discussed further. The pre-trained ResNet50 classification model results achieved at 20 epochs were 100% accuracy and 00% loss, but exhibited a lot of fluctuations in the accuracy graph at almost every alternate epoch, as shown in Figure 6(a). The loss graph Figure 6(b) shows the heavy rise in loss value at the 15th epoch. To improve these results the

learning rate is scheduled and the first 150 layers of the pre-trained model were frozen for 50 epochs only [21]. The pre-trained weights are retained by freezing the early layers, which usually contain the generic features that are already learned by the model. Also with learning rate scheduling, you can begin at a higher rate and progressively lower it as the training goes on for fine-tuning.

The results acquired are shown in Figures 7(a) and 7(b) have maintained a cent per cent of accuracy and negligible loss but still exhibit a sudden fall in accuracy at around the 25th epoch and a sudden rise in loss at the same epoch. To overcome this lacuna the epochs were increased to 150 with the number of batch sizes increased to 26 from 16. The results obtained from this fine-tuning, as shown in Figures 8(a) and 8(b) can be observed that the fall in accuracy is inelible, and also the spike or the sudden increase in loss is illuminated.

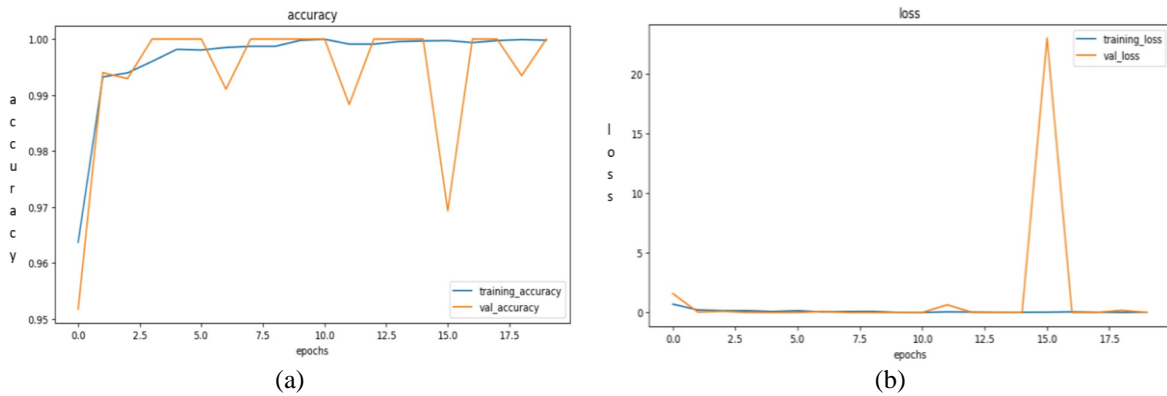


Figure 6. Graph of (a) accuracy and (b) loss at 20 epochs of pre-trained ResNet50 model

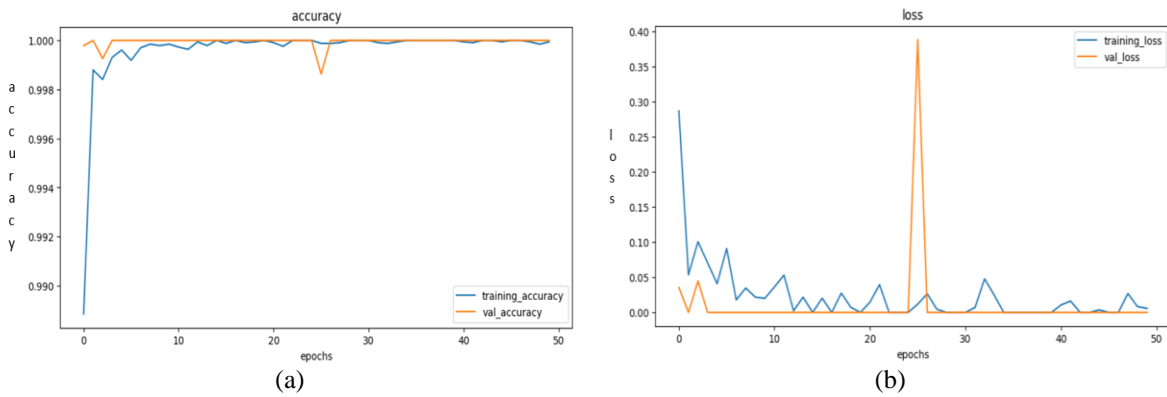


Figure 7. Graph of (a) accuracy and (b) loss at frozen layer and scheduled learning rate at 50 epochs

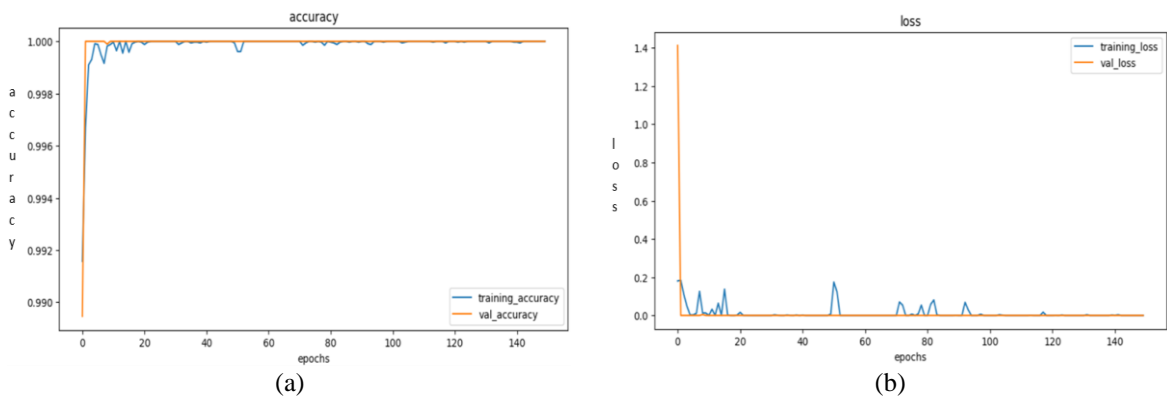


Figure 8. Graph of (a) accuracy and (b) loss at 26 batch size and 150 epoch of ResNet50 pre-trained model

The greater generalization performance can result from a type of regularization that larger batch sizes can offer. To minimize the loss function, a neural network's parameters (weights and biases) must be adjusted during training. By increasing the number of epochs, accuracy may be increased as the model has more time to converge to a better solution. The model requires more time to learn complex patterns and representations in the case of the data sets having complex patterns and representations. The early termination of the training process may not allow, the model to fully capture the underlying patterns in the data. The observation of the accuracy graph shows that 100% accuracy is achieved from the very first epoch itself where ideally there should be exponential growth in the accuracy and exponential fall in the loss as well. Considering that the model could be overfitting the data size was reduced to half and the accuracy for it on the same model and similarly tuned hyperparameters was achieved which is shown in Figures 9(a) and 9(b). No significant change was found after reducing the data size, after the appropriate research work was carried out, it was suspected to be due to a data leakage problem. The accuracy and loss plotted after clearing the data leakage with all the hyperparameters fixed as earlier drastically fell to 79.34% and 19.99% respectively.

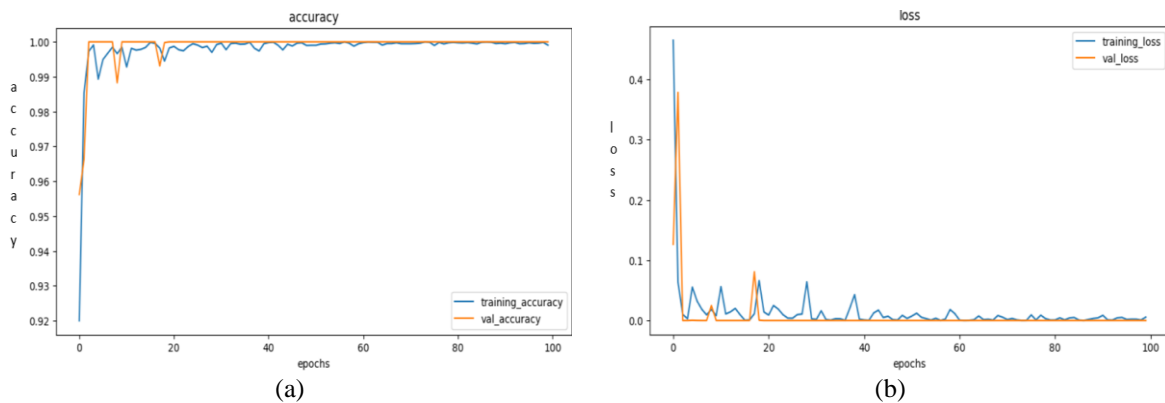


Figure 9. Graph of (a) accuracy and (b) loss at half a dataset size, all hyperparameters fixed, same as earlier

To improve the accuracy of these refreshed 47,500 DR-OCTA images, at first, all the 150 pre-trained ResNet50 layers were unfrozen to use all the pre-trained weights of the model. Then the classification head was removed and replaced by the new classification head with fully connected dense layers with 1,024 neurons. The large number of units has the potential to capture the complex patterns in the images and handle large amounts of data. Further to fine-tune the model's weight the learning rate is scheduled to 0.001 for epoch >40 and 0.005 for epoch >20 [22]. An Adam optimizer is used as it converges faster due to its adaptiveness, batch size is increased to 60 giving more stable updates to the model weights [23], and the accuracy achieved by this fine-tuning of hyperparameters is observed in Figures 10(a) and 10(b).

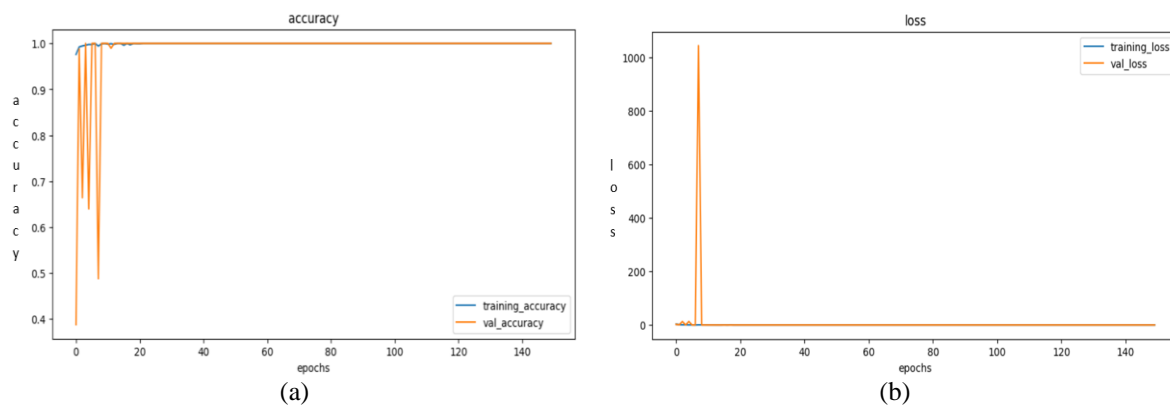


Figure 10. Graph of (a) accuracy and (b) loss of the hyperparameter-tuned ResNet50 pre-trained model

It is observed from the Figure 10 that in the initial 10 epochs there are a lot of fluctuations in the exponential rise of accuracy. In an attempt to reduce these spikes by keeping all the earlier hyperparameters fixed, adding a drop-out layer with a rate of 0.5 means 50% of random outputs are set to zero at each update, thus the model is encouraged for learning more generalized features [24]. The weight decay is used along with the Adam optimizer which helps discourage the weights from becoming too large this regularizes the model [25]. The accuracy and loss after this hyperparameter tuning are shown in Figures 11(a) and 11(b). The reduction in the spikes at the initial epochs is observed. Diabetic retinopathy with five categories, 0-no-DR, 1-mild NPDR, 2-moderate NPDR, 3-sever NPDR, 4-PDR, are very accurately categorized by the model as observed in Figure 12(a) to 12(e).

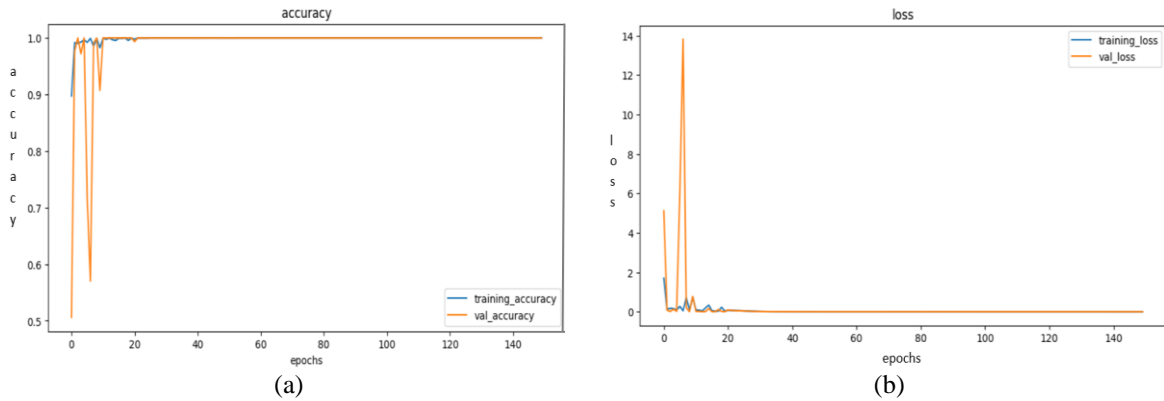


Figure 11. Graph of (a) accuracy and (b) loss of the hyperparameter-tuned ResNet50 with a dropout layer

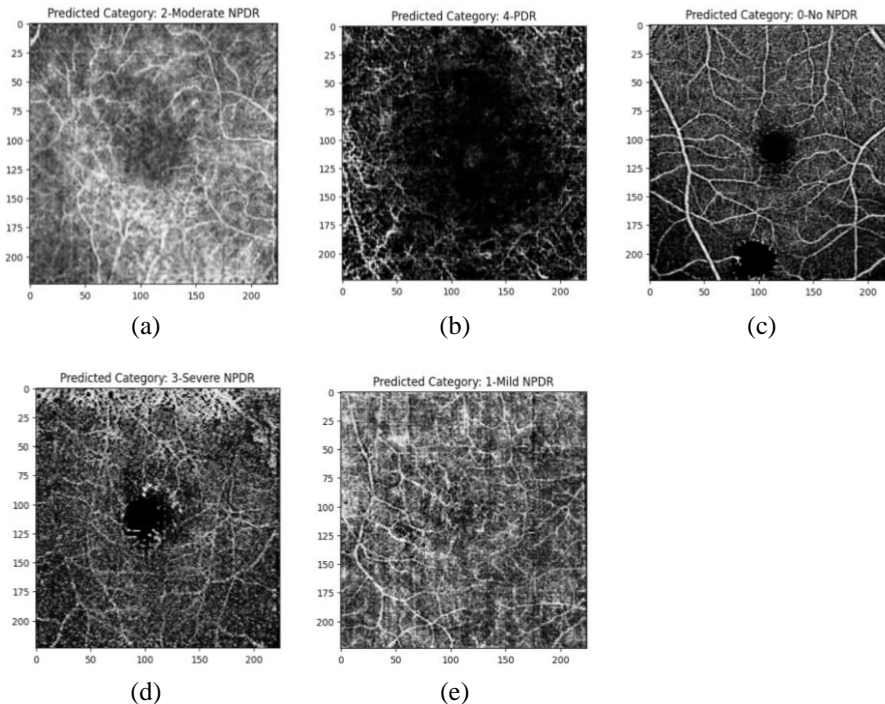


Figure 12. Predicted outputs by pre-trained ResNet50 model, (a) 2-moderate NPDR, (b) 4-PDR, (c) 0-no-NPDR, (d) 3-severe NPDR, and (e) 1-mild NPDR

4. CONCLUSION

In this study, the WGAN model is used for data augmentation over normal augmentation techniques to maintain the quality of the complex DR-OCTA images. The PNG image generation model was further

used for the generation of the images in PNG format from the .h5 file generated by the WGAN model. The ResNet50 model is used for the classification and prediction of the disease with some intense modifications in the hyperparameters. The hyperparameter tuning done in the model was freezing its 150 layers, adding the custom layers, and scheduling the learning rate, which yielded 100% accuracy and nearly 00% loss with a lot of fluctuations. To eliminate this deformity in accuracy the number of epochs was increased which led to a reduction in fluctuations but the accuracy achieved was 100 from the very initial epochs. The research carried out to address this issue led to the conclusion that data leakage was the reason behind it. The resolution of data leakage made the accuracy fall to 75%. Various hyperparameter tuning tasks like unfreezing all layers, fine-tuning the last 30 layers of the model, rescheduling the learning rate adding the Adam optimizer, drop-out layer, and weight decay layers led to better results. The results obtained at various modification stages of the model are discussed in the results in detail. The prediction model correctly predicts the DR-OCTA image disease categories.

This research offers a useful classification and prediction system for diabetic retinopathy, as it analyzes DR-OCTA images quickly and accurately, that is of great assistance to ophthalmologists. Patients will benefit greatly from its ability to detect the disease at an early stage, possibly preventing serious vision problems or blindness. Ophthalmologists will find it easier to begin treatment promptly with this system.




REFERENCES

- [1] R. Cheloni, S. A. Gandolfi, C. Signorelli, and A. Odone, "Global prevalence of diabetic retinopathy: protocol for a systematic review and meta-analysis," *BMJ Open*, vol. 9, no. 3, Mar. 2019, doi: 10.1136/bmjopen-2018-022188.
- [2] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, 2017, pp. 214–223.
- [3] X. Yuan *et al.*, "Image enhancement of wide-field retinal optical coherence tomography angiography by super-resolution angiogram reconstruction generative adversarial network," *Biomedical Signal Processing and Control*, vol. 78, Sep. 2022, doi: 10.1016/j.bspc.2022.103957.
- [4] J. Lyu, Y. Zhang, Y. Huang, L. Lin, P. Cheng, and X. Tang, "AADG: automatic augmentation for domain generalization on retinal image segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3699–3711, Dec. 2022, doi: 10.1109/TMI.2022.3193146.
- [5] A. Chinkamol *et al.*, "OCTAve: 2D en face optical coherence tomography angiography vessel segmentation in weakly-supervised learning with locality augmentation," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 6, pp. 1931–1942, Jun. 2023, doi: 10.1109/TBME.2022.3232102.
- [6] C.-H. Hua *et al.*, "Convolutional network with twofold feature augmentation for diabetic retinopathy recognition from multi-modal images," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2686–2697, Jul. 2021, doi: 10.1109/JBHI.2020.3041848.
- [7] S. Zhou, D. Yu, Y. Cai, Y. Zhang, B. Li, and W. Li, "TCAM-Resnet: a convolutional neural network for screening DR and AMD based on OCT images," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2022, pp. 1830–1835, doi: 10.1109/BIBM55620.2022.9995218.
- [8] K. A. Alavee *et al.*, "Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence," *IEEE Access*, vol. 12, pp. 73950–73969, 2024, doi: 10.1109/ACCESS.2024.3405570.
- [9] A. Jabbar *et al.*, "Alision-based diabetic retinopathy detection through hybrid deep learning model," *IEEE Access*, vol. 12, pp. 40019–40036, 2024, doi: 10.1109/ACCESS.2024.3373467.
- [10] Z. Wang, S. Chen, T. Liu, and B. Yao, "Multi-branching temporal convolutional network with tensor data completion for diabetic retinopathy prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 3, pp. 1704–1715, Mar. 2024, doi: 10.1109/JBHI.2024.3351949.
- [11] B. N. Jagadesh, M. G. Karthik, D. Siri, S. K. K. Shareef, S. V. Mantena, and R. Vatambeti, "Segmentation using the IC2T model and classification of diabetic retinopathy using the rock hyrax swaram-based coordination attention mechanism," *IEEE Access*, vol. 11, pp. 124441–124458, 2023, doi: 10.1109/ACCESS.2023.3330436.
- [12] Q. Li *et al.*, "Diagnosing diabetic retinopathy in OCTA images based on multilevel information fusion using deep learning framework," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–10, Aug. 2022, doi: 10.1155/2022/4316507.
- [13] D. Le *et al.*, "Transfer learning for automated OCTA detection of diabetic retinopathy," *Translational Vision Science & Technology*, vol. 9, no. 2, Jul. 2020, doi: 10.1167/tvst.9.2.35.
- [14] G. Ryu, K. Lee, D. Park, S. H. Park, and M. Sagong, "A deep learning model for identifying diabetic retinopathy using optical coherence tomography angiography," *Scientific Reports*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-02479-6.
- [15] P. Zang *et al.*, "DcardNet: diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 6, pp. 1859–1870, Jun. 2021, doi: 10.1109/TBME.2020.3027231.
- [16] J. Zhu, G. Yang, and P. Lio, "How can we make GAN perform better in single medical image super-resolution? a lesion-focused multi-scale approach," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Apr. 2019, pp. 1669–1673, doi: 10.1109/ISBI.2019.8759517.
- [17] A. Manassakorn *et al.*, "GlauNet: glaucoma diagnosis for OCTA imaging using a new CNN architecture," *IEEE Access*, vol. 10, pp. 95613–95622, 2022, doi: 10.1109/ACCESS.2022.3204029.
- [18] M. Khalid, J. Baber, M. K. Kasi, M. Bakhtyar, V. Devi, and N. Sheikh, "Empirical evaluation of rectified activations in convolution network," in *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*, Jul. 2020, pp. 204–207, doi: 10.1109/TSP49548.2020.9163446.
- [19] A. G. Devi *et al.*, "Implementing ResNet-50 transfer learning model for diagnosing OCT images for detecting and classifying DME abnormalities," *Journal of Theoretical and Applied Information Technology*, vol. 15, no. 15, 2023.
- [20] E. Rezende, G. Ruppert, T. Carvalho, F. Ramos, and P. de Geus, "Malicious software classification using transfer learning of ResNet-50 deep neural network," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2017, pp. 1011–1014, doi: 10.1109/ICMLA.2017.00-19.




- [21] Y. Jin *et al.*, “Autolrs: automatic learning-rate schedule by bayesian optimization on the fly,” *arXiv preprint arXiv:2105.10762*, 2021.
- [22] Y. Huang, L. Lin, P. Cheng, J. Lyu, R. Tam, and X. Tang, “Identifying the key components in ResNet-50 for diabetic retinopathy grading from fundus images: a systematic investigation,” *Diagnostics*, vol. 13, no. 10, May 2023, doi: 10.3390/diagnostics13101664.
- [23] S. L. Smith, P. J. Kindermans, C. Ying, and Q. V. Le, “Don’t decay the learning rate, increase the batch size,” *arXiv preprint arXiv:1711.00489*, 2017.
- [24] M. K. Yaqoob, S. F. Ali, I. Kareem, and M. M. Fraz, “Feature-based optimized deep residual network architecture for diabetic retinopathy detection,” in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, Nov. 2020, pp. 1–6, doi: 10.1109/INMIC50486.2020.9318096.
- [25] M. Rao, M. Zhu, and T. Wang, “Conversion and implementation of state-of-the-art deep learning algorithms for the classification of diabetic retinopathy,” *arXiv preprint arXiv:2010.11692*, 2020.

BIOGRAPHIES OF AUTHORS



Pranali Pradeep Hatode    received a B.Eng. degree in electronics engineering from Rashtrasanta Tukadoji Maharaj Nagpur University and a M.Eng. degree in electronics and telecommunication engineering from the University of Mumbai. Currently, she is an assistant professor at K J Somaiya Institute of Technology, in the Department of Electronics and Telecommunication. She has 2 Published at the 5th IEEE International Conference on Advances in Science and Technology 2022 (5th IEEE-ICAST) Publications at IEEE Xplore, Scopus Indexed, and has a patent in process on “Braille Key Board Printer”, application No. E-12/2633/2022/MUM. Her research interests are in image processing, implementation of AI techniques in medical imaging, AI, and ML in healthcare. She can be contacted at phatode@somaiya.edu.



Maniroja Edinburgh    holds a Ph.D. in electronics from Sant Gadge Baba Amravati University, Maharashtra, India. She is the Head of the Department of Artificial Intelligence and Data Science at Thadomal Shanhi Engineering College, Mumbai, India. She holds 31 years of work experience in academics and is a recognized postgraduate teacher and a Ph.D. Guide of the University of Mumbai. She is a member of the syllabus revision committee, a vice-chancellor nominee as a subject expert for UGC interviews, a member of a minor research grant committee in the Faculty of Engineering and Technology, a local inquiry committee member at the University of Mumbai, an examination panel member at various universities and a member subject specialist for the Maharashtra Public Service Commission. Apart from this, she has delivered lectures on various cutting-edge technology topics at various colleges and universities. She received the outstanding reviewer award from Elsevier Journal “Electronics and Computer Engineering” in Jan 2016, has around 30 national/international journal publications, around 24 national/international conferences and received a minor research grant from the University of Mumbai for the research projects “Touch in digital camera” and “RF controlled environmental cleaning project”. She can be contacted at maniroja@thadomal.org.