International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.14, No.1 (Jul-23)

http://dx.doi.org/10.12785/ijcds/140111

Transfer Learning based Diabetic Retinopathy Classification

S Charulatha[1], Dharani C[1] and Suguna M[1]

[1]*Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai, India*

Received 14 Jun. 2022, Revised 6 May. 2023, Accepted 21 May. 2023, Published 1 Jul. 2023

Abstract: Diabetic retinopathy(DR), a disease that affects the eye by causing damage to the blood vessels of the light sensitive retina. Diabetic retinopathy at the earlier stage might cause no or very minimal vision issues but if not cured it might lead to loss of vision as well. Diabetic retinopathy, particularly in people of working age, is the main cause of blindness among diabetic patients in developing countries. Normal, mild, moderate, severe, and PDR (Proliferative Diabetic Retinopathy) are the five stages of diabetic retinopathy.. Usually, trained experts look into the colored fundus image of the eye to examine the stage of DR. But this process is manual and is highly time consuming. It could also sometimes lead to error-prone results. Therefore, various Computer Vision approaches are proposed to detect the stage of DR in the fundus image, but they lack in particular classifying the DR at lower stages. The proposed methodology compares various CNN models (Densenet201, InceptionV3, Resnet50) along with processing specifically to extract maximum features from fundus image to classify the stage of DR with the publicly available kaggle dataset. The proposed methodology compares the algorithms based on Quadratic Kappa Score where the obtained score for Densenet201, InceptionV3 and Resnet50 are 0.931,0.956 and 0.926 respectively.

Keywords: Machine Learning , CNN, Diabetic Retinopathy, Densenet201, Resnet50, InceptionV3, Computer Vision.

1. INTRODUCTION

A medical condition called Diabetic Retinopathy, also referred to as diabetic eye disease, is caused by diabetes mellitus and damages the retina. It may develop as a result of having high blood sugar levels, which can harm blood vessels all over the body, including those in the retina. As was mentioned in the previous section, blindness is a significant issue in developing nations. The five stages of DR include normal, mild, moderate, severe, and proliferate. While the micro-aneurysm (MR) may appear to be a small red dot in the earlier stages, as the condition progresses to the later stages, the MA deepens and develops into a hemorrhage that resembles a flame, which results in blindness.

By 2025, there are anticipated to be 592 million patients with DR, up from 382 million currently according to [1]. DR detection is regarded as one of the more difficult diseases to diagnose because by the time human examiners submit their reports, the disease will have advanced further. DR detection is currently done manually, which takes a lot of time and resources and makes the process slow. Therefore, it is necessary to put in place an automated system that would identify and categorize the stage of DR in the human eye in order to lessen this boilerplate.

The DR is currently categorized using methods of image processing. There are also methods that classify DR or No DR using machine learning algorithms. However, the proposed methodology uses three different types of transfer learning models, hyper parameter tuning, and DR-specific pre-processing to categorize the stage of DR into the five classes that were covered in the earlier section.

2. RELATED WORKS

Multitasking approach for diabetic retinopathy detection has been employed[2].This paper uses a regression model and classification model for detection and classification of DR.Classification model will help finding the difference in characteristics between five stages of DR whereas the regression model will find the inter-dependent characteristics between the stages.The results of both models are concatenated and the final classification of class is determined.The paper proposed two models in multitasking approach of SEDensenet and Xception.The maximum Kappa score for APTOS dataset was shown by the Xception model which is 0.90.

E-mail address: charulathas@student.tce.edu, dharanic@student.tce.edu, mscse@tce.edu

The VGG-NiN model[3] proposed an architecture for DR classification which uses less possible learnable parameters for fast convergence of the model and less training time.The proposed model stacked the VGG16 SPP layer and Network-in-Network which makes the model learn nonlinear features and make it better in terms of accuracy.

This article explains a transfer learning-based technique for neovascularization detection[4]. Using four pre-trained Convolutional Neural Network (CNN) models, including AlexNet, GoogLeNet, ResNet18, DenseNet201, and InceptionV3, the effectiveness of the transfer learning method is examined.

This particular paper[5] helped us in exploring on how Convolutional Neural Network is employed in classifying the stage of diabetic retinopathy based on symmetric convolutional neural network.This paper also gave us an insight on using layers of CNN on predicting the minute points causing blindness.

In order to better detect lesions in the fundus image, the hybrid CNN-SVD model[6] for dimensionality reduction and Extreme Learning Machine(ELM) Algorithm for classification used preprocessing methods like BenGraham's principle and CLAHE.

In[7], a deep CNN is used. The deep CNN uses symmetric convolutional structures to identify various types of lesions in images of the diabetic retina. Convolution, maximum pooling, and average pooling layers are used in the implemented approach for feature filtering, and the outcomes with the three features are compared. This implementation helped us in deciding upon which pooling layers to be added to the model. The implemented approach concluded that average pooling performs better in terms of identifying the tiny lesions from the fundus image with an accuracy of 0.936 whereas max pooling and convolution was comparatively lower with 0.920 and 0.932 respectively.

In[8], This paper used GoogleNet to classify the DR images with a cloud background for prediction. The authors claim that the system they proposed reduces the time in classification of DR stages whereas there was no explanation on individual class wise performance.

As a result, from the background literature reviews, it is evident that though the existing implementations gave us solid understanding of various methodologies to approach diabetic retinopathy detection, but also lacks in various areas like the complexity of the model, Cohen's kappa score. In the proposed methodology, few of the issues are addressed along with pre-processing specific to the retinal image for the selected transfer learning models InceptionV3, Densenet201 and Resnet50 respectively.

3. PROPOSED WORK

A. *Dataset Description*

The dataset used is the Kaggle APTOS 2019 dataset.It consists of labeled fundus images of five classes namely 0-No DR,1-Mild,2-Moderate,3-Severe,4-Proliferative DR.The dataset has 3662 training images. The train and test images are split on the basis of 80- 20 proportions. Figures 1 and 2 shows the distribution of images in each class and the representation of each stage of DR.

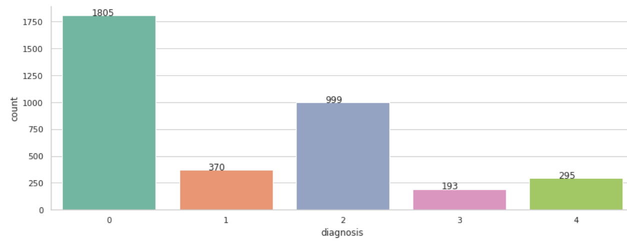

Figure 1. Dataset distribution

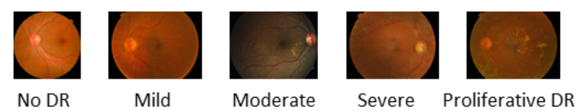

Figure 2. Stages of DR

B. *Preliminaries*

1) *Convolutional Neural Networks*

Convolutional Neural networks(CNN) are used in deep learning models to extract complex features in images.The features can be identified as the simplest part to identify an image and classify. For e.g., detecting eyes of a human in an image which is a feature of human image.CNNs consist of many layers of feed forward network where each of the layers detect simple to complex features from the input. Convolutional Neural Network has a series of layers such as the convolution layer, activation, and pooling layer for extracting features and finally there will be a fully connected layer to classify images.

The Convolution layer has filters which are square matrices which will slide over the input matrix and a dot product is performed. Generally the image input will be expressed as ($ht \times wt \times 3$).The resulting matrix is called the activation map. On this activation map, an activation function is applied. Activation function introduces non-linearity to the CNN.ReLU activation is used mostly in CNNs which activates the input if it is positive. Leaky ReLU, an activation which is a modified version of ReLU, is also used in CNNs.

To the output feature map after applying activation, pooling operation is performed. Generally, there are two types of pooling, Max pooling and Average pooling. The purpose of pooling is to reduce the size of the feature maps to make the computation fast and simpler. The pooling works the same way as the convolution layer by sliding over the feature map but it will take the max(or average) value from the region being overlapped by the kernel unlike dot product. One more advantage of pooling operation is it reduces noise in an image as it takes the max value only.

2) *Transfer Learning*

Transfer learning is a technique that uses large pre-trained models for similar related tasks.Transfer learning is a method where a large pretrained model is used in a similar related task. This is basically, instead of training a model from scratch, the pretrained model weights are used where it has seen a lot of similar data. The main advantage of transfer learning is the training time where training a model from scratch takes a lot of time and computation power. Also, the amount of data needed for training a model is huge which in most cases is not available. Most of the pretrained models are trained with the Imagenet dataset which has over 1000 classes and around 1 million training images.

a) *Densenet201*

Densenet 201[9] is a convolutional neural network and it is 201 layers deep. In Dense Convolutional Network(DenseNet), each layer is connected to every other layer as shown in Figure 3 where in other networks each layer will connect only to the subsequent layers.Therefore the number of layers in a densenet is given by $L(L+1)/2$. The advantage of DenseNet is that it requires fewer parameters while avoiding redundant feature map training. Another advantage of DenseNets is that gradient flows make training easier because they are directly linked to the previous layer.

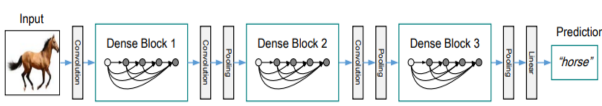

Figure 3. Dense blocks in the DenseNet architecture with transition layers between the dense blocks (*source:[9]*)

b) *Resnet50*

Resnet models[10] are built by the intuition of skip connections. Deeper networks can detect complex features but training them comes with a problem of vanishing gradient which results in saturation of accuracy. To overcome this problem, the Resnet comes with a solution of Residual blocks where there is a direct connection that skips some layers in between Figure 4. In this way the vanishing gradient problem is handled where there will be an alternate path to the gradient flow.

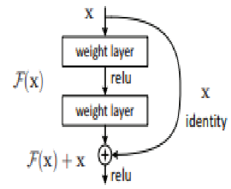

Figure 4. The identity mapping between layers in the Resnet architecture (*source:[10]*)

c) *InceptionV3*

Inception models were developed by Google. The deep models tend to overfit the data, so these inception models have different sized filters at the same level so this keeps the network wider rather than deeper and also avoids overfitting. The Inception v3[11] is an improved version where the larger convolutions were simplified into smaller convolutions making the model less computational Figure 5.

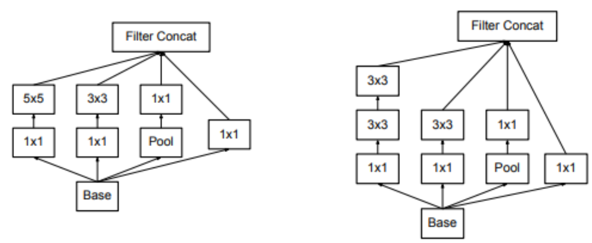

Figure 5. The left one shows the larger convolutions involved in the previous Inception models while the right one replaces those with simpler convolutions in InceptionV3.(*source:[11]*)

C. *Proposed work*

In the proposed work as shown in Figure 6, the dataset is first pre-processed to gray-scale. Then in order to focus only on the circular part of the eye, the image is then circular cropped.This crop was used to remove black areas from original square images, and also crop a circular area based on the image center.Then the image is resized to 224 * 224 for better feature extraction. After the per-processing, pre-trained model weights of Densenet201, InceptionV3 and Resnet50 are loaded by removing the top layer. The layers are altered with Global Average Pooling which will extract the fine details from the fundus image. In order to avoid overfitting, dropouts are performed. Then the model is optimized using adam optimizer and then with SGD. Finally, the performance of the transfer learning models with the proposed approach is compared against Cohen's kappa score.

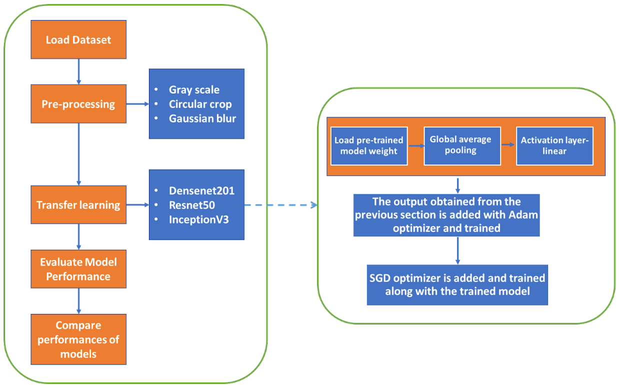

Figure 6. System design of proposed work

4. METHODOLOGY

A. Pre-processing

The input image undergoes the following image pre-processing techniques which are listed below:

1) Image Resizing: To facilitate preprocessing in later stages, the image is resized to $224 \times 224px$ width.
2) Grayscale Conversion: The RGB-formatted input image that we typically receive has a lot of details that are not necessary for our process. The values will be decreased from 256 to a smaller value as a result of converting the image from RGB to grayscale.
3) Noise removal: Image noises are imperfections in the image that result from bad photographs.
4) There are many types of noise, including salt and pepper noise and Gaussian noise. In order to remove the noise from the image, we used Gaussian blur.
5) Circular crop: Since the lesions in the fundus image are the primary focus of diabetic retinopathy detection, circular cropping is used.

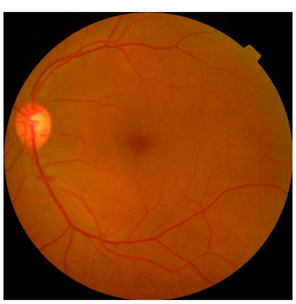

Figure 7. Image before pre-processing

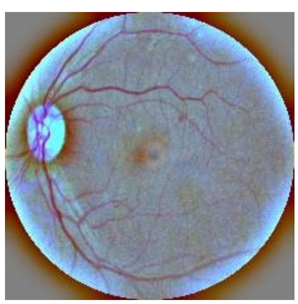

Figure 8. Image after pre-processing

After this, the image is passed to the transfer learning models which detects the stage of the retinopathy from 0 to 4.

B. Model implementation

In the proposed system the pre-trained model weights of Densenet201, InceptionV3 and Resnet50 are loaded respectively where the top layers of the model are removed. Global Average pooling is used in order to average the intensity values which captured fine details from the fundus image.The image dataset is augmented on the fly using keras data generator which the images are rescaled by dividing by dividing it by 255 and then it is randomly rotated by 360 degrees. All these functions are done on the fly which would reduce space complexity of the overall model. Then in order for the model to avoid overfitting drop outs are performed .Later, each model was trained with an optimizer. Optimizers are used because it alters the weights values accordingly which increases the model accuracy. Firstly the model is trained using Adam optimizer, it is chosen because Adam optimizers are computationally less expensive. But while executing the model with Adam optimizer the validation loss was greater than the accuracy. Later the model was fine tuned with SGD optimizer which converges earlier with a minimum loss than Adam optimizer. Therefore at the end of SGD optimizer the loss was minimal and the accuracy was better. Then the results obtained from Adam and SGD are combined. When combined the accuracy was furthermore improved as the Adam and SGD are exploited at the same time by taking the best of two.

C. Performance metric

Since the dataset is imbalanced, accuracy will not be a good metric for evaluating the performance of the model because the accuracy will be more biased towards the class with more data.So, here we use Quadratic weighted kappa score for evaluating the performance of the model. Cohen's kappa score is calculated for each model using the following equation.

$$K = \frac{P_o - P_e}{1 - P_e}$$

Where

P_o: Observed probability

P_e: Expected probability

Cohen's kappa score values range from -1 to 1. As the Cohen's kappa score converges to 1 it means that the model has predicted the stage of the DR correctly.

5. RESULTS AND DISCUSSIONS

We have implemented and compared three transfer learning models and obtained the following results.

A. Densenet201

The Densenet201 model has train and val accuracy in the range 0.80 to 0.85.Since diabetic retinopathy detection falls under the category of classification problem, it is appropriate to visualize the results in the form of heatmap.The Densenet201 model was best in detection the first level of DR than the other levels in the train and val set. From the result visualizations Figures 9 and 10, we can infer that the classes 3 and 4 are not performing well. The kappa score of DenseNet is 0.931

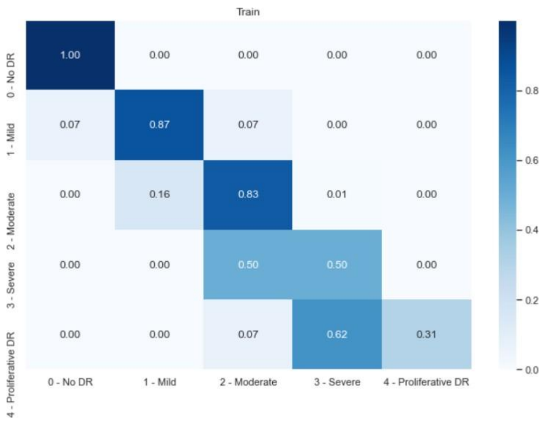

Figure 9. Densenet201-Heatmap of Train set

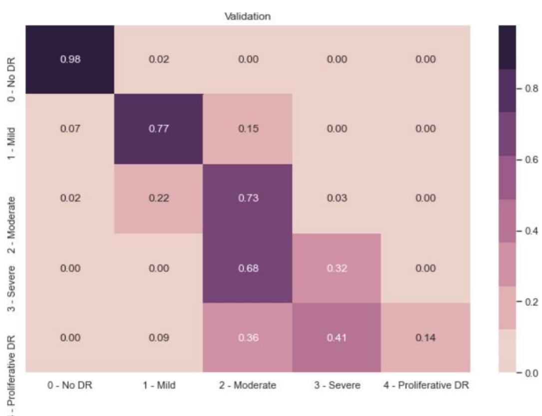

Figure 10. Densenet201-Heatmap of Val set

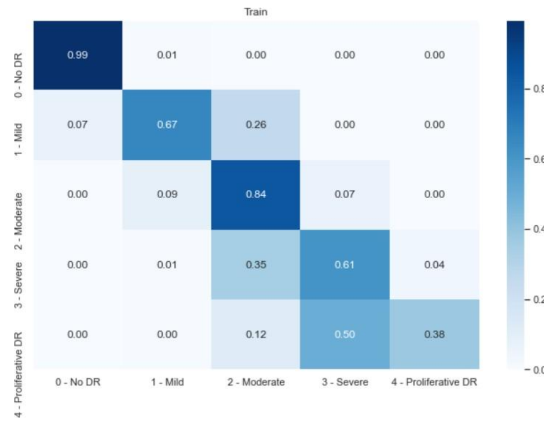

Figure 11. Resnet50-Heatmap of Train set

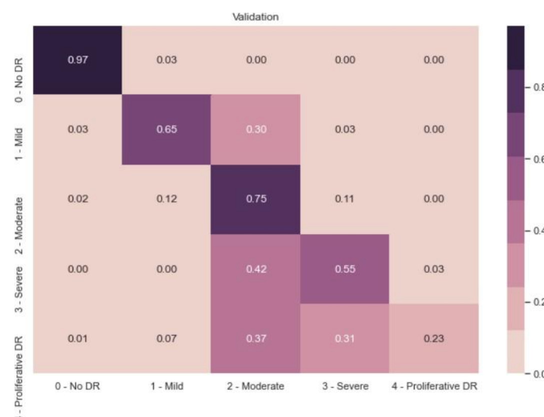

Figure 12. Resnet50-Heatmap of Val set

B. *Resnet50*

The Resnet50 model achieved a train and val accuracy of about 0.75 to 0.80 which is relatively lower than the Densenet201. From the heatmap as shown in Figures 11 and 12, it can be seen that the Resnet50 model was able to recognize class 0 and class 2 better than all other types.But it shows poor recognition for class 5.So, we cannot conclude that the model will predict all classes correctly. The model is more biased towards classes 0 and 2. The model was able to achieve a kappa score of 0.926

C. *InceptionV3*

Inception V3 accuracy was higher than Densenet201 and Resnet50 models whose train and val accuracy was about 85%. From the heatmap visualization shown in Figures 13 and 14, we can see that the model was able to identify type 0, 1 2 and 3 which is higher than the other two models. Especially it is performing better in classifying the proliferative DR stage which if wrongly predicted leads to blindness. Also the cohen's kappa score is 0.956 which is better than the other two models.

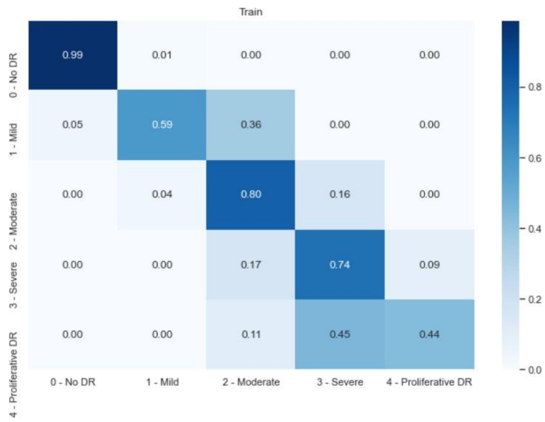

Figure 13. InceptionV3-Heatmap of Train set

Figure 14. InceptionV3-Heatmap of Val set

From the results of the experiments with the three models, it can be inferred that the InceptionV3 model with 0.95 Kappa score has outperformed Resnet50 and DenseNet201, also outperforms the existing Multitasking model[2] which has maximum Kappa score of 0.88. Table I shows the comparison of these models.

Models	Kappa score
Densenet201	0.931
Resnet50	0.926
InceptionV3	0.956
Multitasking Deep Learning Model for Detection of Five Stages of Diabetic Retinopathy[2]	0.88

TABLE I. Comparison of Kappa scores of models

6. CONCLUSION AND FUTURE WORK

The proposed approach of transfer learning in diabetic retinopathy detection with three models Densenet201, Resnet50 and InceptionV3 proved to be better than existing approaches and among all the three InceptionV3 outperformed better than the other two in terms of Cohen's Kappa Score for the Kaggle dataset. The existing Multitasking model[2] achieved a kappa score of 0.88 while the proposed approach with Densenet201, Resnet50 and InceptionV3 models achieved a kappa score of 0.931, 0.926 and 0.956 respectively. The kappa score of the proposed approach is higher than the existing approach while among the proposed models, InceptionV3 outperformed densenet201 and resnet50. In the future we would like to work on further fine tuning in order to increase the model performance even more as well as train a few more models for detecting diabetic retinopathy among people.

References

[1] M. Nahiduzzaman, M. Robiul Islam, M. Omaer Faruq Goni, M. Shamim Anower, M. Ahsan, J. Haider, and M. Kowalski, "Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and elm classifier," *Expert Systems with Applications*, vol. 217, p. 119557, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423000581

[2] S. Majumder and N. Kehtarnavaz, "Multitasking deep learning model for detection of five stages of diabetic retinopathy," *IEEE Access*, vol. 9, pp. 123 220–123 230, 2021.

[3] Z. Khan, F. G. Khan, A. Khan, Z. U. Rehman, S. Shah, S. Qummar, F. Ali, and S. Pack, "Diabetic retinopathy detection using vgg-nin a deep learning architecture," *IEEE Access*, vol. 9, pp. 61 408–61 416, 2021.

[4] M. C. S. Tang, S. S. Teoh, H. Ibrahim, and Z. Embong, "A deep learning approach for the detection of neovascularization in fundus images using transfer learning," *IEEE Access*, vol. 10, pp. 20 247–20 258, 2022.

[5] B. Yang, T. Li, H. Xie, Y. Liao, and Y.-P. P. Chen, "Classification of diabetic retinopathy severity based on gca attention mechanism," *IEEE Access*, vol. 10, pp. 2729–2739, 2022.

[6] M. Nahiduzzaman, M. R. Islam, S. M. R. Islam, M. O. F. Goni, M. S. Anower, and K.-S. Kwak, "Hybrid cnn-svd based prominent feature extraction and selection for grading diabetic retinopathy using extreme learning machine algorithm," *IEEE Access*, vol. 9, pp. 152 261–152 274, 2021.

[7] T. Liu, Y. Chen, H. Shen, R. Zhou, M. Zhang, T. Liu, and J. Liu, "A novel diabetic retinopathy detection approach based on deep symmetric convolutional neural network," *IEEE Access*, vol. 9, pp. 160 552–160 558, 2021.

[8] D. Pruthviraja, A. B. C., and S. C. N., "Efficient Local Cloud-Based Solution for Diabetic Retinopathy Detection," *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, vol. 16, no. 3, pp. 39–46, May 2021.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

S Charulatha S Charulatha has done her Bachelor of Engineering in Computer Science and Engineering at Thiagarajar College of Engineering. Her area of interest lies in Computer Vision and Image processing and has contributed to technical events promoting Open CV.

Dharani C Dharani has completed her Bachelor of Engineering in Computer Science and Engineering at Thiagarajar College of Engineering. Dharani has a keen interest towards problem solving and have actively worked on Computer Vision related projects.

Suguna M M. Suguna is presently working as an Assistant Professor at Department of Computer Science and Engineering, Thiagarajar Col- lege of Engineering, TamilNadu, India. She received her B.E. and M.E. degree in Computer Science and Engineering in 2004 and 2008 respectively. She received her Ph.D. in Information and Communication Engineering in 2019 from Anna Univer- sity, Chennai, TamilNadu, India. Her research interests include Applied Machine Learning, Computer Vision, Smart Nano devices and Applied Cryptography.