A Deep Learning based Hybrid Model for Classification of Diabetic Retinopathy

Sandeep Madarapu
Department of EC
NIT Rourkela
Odisha, India
madarapusandeep@gmail.com

Samit Ari
Department of EC
NIT Rourkela
Odisha, India
samit@nitrkl.ac.in

Kamalakanta Mahapatra
Department of EC
NIT Rourkela
Odisha, India
kkm@nitrkl.ac.in

Abstract—**Diabetic retinopathy (DR) is a complication of diabetes that effects the retina. Swelling of the retinal blood vessels due to excessive sugar in diabetic patients can cause DR, which damages the retina. Diagnosis of DR is tedious and time-consuming for clinical experts, hence a computer-aided-diagnosis (CAD) tool is required to detect DR automatically. This paper proposes a novel method for detecting multi-class DR using the Xception model and random forest. First, the fundus images are pre-processed with the contrast-limited adaptive histogram equalization (CLAHE) technique to enhance image contrast by removing the embedded noise. This work proposes a hybrid deep convolutional neural network (DCNN) that concatenates the extracted features from various layers of the pre-trained Xception model to improve the performance. The dimensions of the local features are reduced using linear discriminant analysis (LDA). The resulting features are concatenated and utilized for training the random forest for DR classification. The performance of the proposed method is validated on the publicly available APTOS-2019 database. The experimental results show that the proposed technique is better compared to the state-of-the-art techniques.**

Index Terms—**CLAHE, Deep learning, Diabetic retinopthy, Fundus images LDA, ROC, Xception.**

I. Introduction

Diabetic retinopathy (DR) is the most significant reason for vision impairment in middle-aged people. It is vital to regularly undergo eye screening to prevent vision loss or blindness. DR is estimated to affect about one-third of all diabetic patients. According to the international diabetes federation (IDF), diabetes affected approximately 460 million people across the world in 2019, with the number predicted to increase to 700 million by 2045 [1]. The retina's tiny blood vessels are damaged by diabetes, causing it to leak. Non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) are the two kinds of diabetic retinopathy. NPDR is the first stage of Diabetic Retinopathy (Stage I) and is categorized into mild, moderate, and severe NPDR [2], [3]. Small micro-aneurysms begin to form in mild NPDR. Micro-aneurysms are pin-point hemorrhages that are very small in size and have the shape of a balloon. In Moderate NPDR, Capillaries burst as the disease progresses, resulting in larger dark red areas (flame-like structure) known as hemorrhages (HA). it also generates hard exudates (HE) by causing capillaries to become more porous. Soft exudates (SE) are pale spots with soft edges that appear when capillary support weakens [1]. In the severe NPDR Stage, blood flow to various regions in the retina is interrupted due to the blockage of blood vessels. As a result, new blood vessels, known as intra microvascular abnormalities (IMRA), cause venous abnormalities like loops and beading [4]. The normal and mild stages have a similar appearance as a result, detecting the Mild stage is a challenging task. As the disease progress, it leads to PDR, which is also stage II of Diabetic Retinopathy. In PDR, new blood vessels begin to form on the beneath surface of the retina, or vitreous is called neovascularization [5], [6], which may lead to loss of eyesight. These newly formed blood vessels are fragile and can easily be broken by releasing blood into the vitreous, a condition known as vitreous hemorrhage [7]. These new blood vessels appear as a rim-like structure. The new blood vessel formation has been sub-divided into two types, neovascularization within the disc region called NVD and neovascularization other than the disk region called NVE. NVD is worse when compared to NVE. It is critical to have regular screenings to avoid DR. However, DR diagnosis and screening must be performed by an ophthalmologist. As the number of patients increases, so does the burden on a limited number of ophthalmologists, making fundus image analysis more difficult. As a result, a computer-aided diagnosis system that can automatically examine fundus images is required, reducing the burden and analyzing time on an ophthalmologist. The main contributions of the proposed work is organised as follows:

- Local and global features are extracted using proposed Xception model, which improves the classification performance.
- linear discriminant analysis (LDA) technique is utilized to reduce the dimensionality of local features, which decreases the training time of the random forest classifier.
- End-to-end system utilized less training data as Machine learning algorithm is incorporated in the methodology for the classification of fundus images.

The rest of the manuscript is organized as follows. Section II describes the proposed methodology for the detection of DR, Section III demonstrates the experimental results and discussion, finally the script is concluded with Section IV.

TABLE I
A DETAILED LITERATURE SURVEY ON DR CLASSIFICATION

Literature	Dataset	Number of classes	Methodology
Islam *et al.*, 2022 [1]	APTOS 2019	2 & 5	Pre-trained Xception transfer learning model with the supervised contrastive learning method is utilized in this work to detect DR.
Shankar *et al.*, 2020 [5]	MESSIDOR	4	Classification of the histogram-based segmented image using hyper-parameter tuning of the Inception-v4 model
Jyostna *et al.*, 2021 [3]	APTOS 2019	2 & 5	Classification of the histogram-based segmented image using hyper-parameter tuning of the Inception-v4 model
P. Saranya *et al.*, 2020 [6]	MESSIDOR	4	After removing the optic disc, canny-edge detection is applied to the fundus image to segment blood vessels and train a convolutional neural network for DR classification.
Mohamed *et al.*, 2022 [8]	APTOS 2019	2 & 5	Convolutional block attention module (CBAM) with DenseNet169 as a backbone encoder is used for feature extraction to classify DR grade.
Harshit *et al.*, 2021 [9]	EyePACS	2 &5	Concatenation of convolution neural networks (CNN)
Grace *et al.*, 2022 [4]	EyePACS & MESSIDOR	4 & 5	The features extracted from VGG16 and Inception-v3 are concatenated by providing CLAHE images and contrast-enhanced canny edge detection images, respectively.
S. Gayathri *et al.*, 2020 [10]	IDRiD & MESSIDOR,	2 & 5	Extraction of the features is accomplished through the utilization of sped-up robust features and binary robust invariant scalable key points. The extracted features are combined and given to machine learning models for classification.
G. Kalyani *et al.*, 2021 [11]	Messidor	4	Convolutional layers and a primary capsule layer are used to extract the features. For classification, the Class capsule layer is used.

II. METHODOLOGY

The database utilised in the proposed work, pre-sprocessing of the fundus images for effective classification, and the proposed deep learning architecture for the fundus image classification is discussed in this methodology section.

A. Dataset

Many of the researchers are using publicly available data sets to assess DR. In this work, the Asia Pacific tele-ophthalmology society (APTOS-2019) public dataset is considered from the Kaggle competition. The fundus images in this dataset are provided by Arvind eye hospital in India and are divided into five stages: normal, mild, moderate, severe, and proliferative DR. The dataset includes 3662 labeled images and 1928 unlabeled images. However, only 3662 images with labels are considered in this study, which are divided into train and test samples in the proportions of 80%, and 20% respectively as shown in Table II.

TABLE II
APTOS-2019 DATASET RETINOPATHY GRADE DISTRIBUTION

DR Stage	Total No.of images	No. of train images	No. of test images
Normal (0)	1805	1438	367
Mild (1)	370	301	69
Moderate (2)	999	798	201
Severe (3)	193	162	31
PDR (4)	295	230	65

B. Preprocessing

The brightness and resolution of the fundus images in our dataset vary greatly because they were produced using various

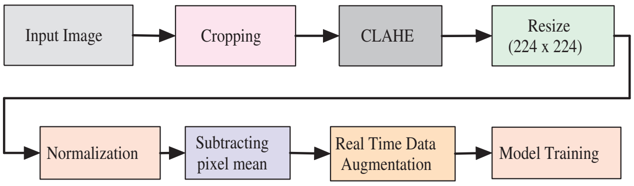

Fig. 1. Pre-processing of fundus images using CLAHE.

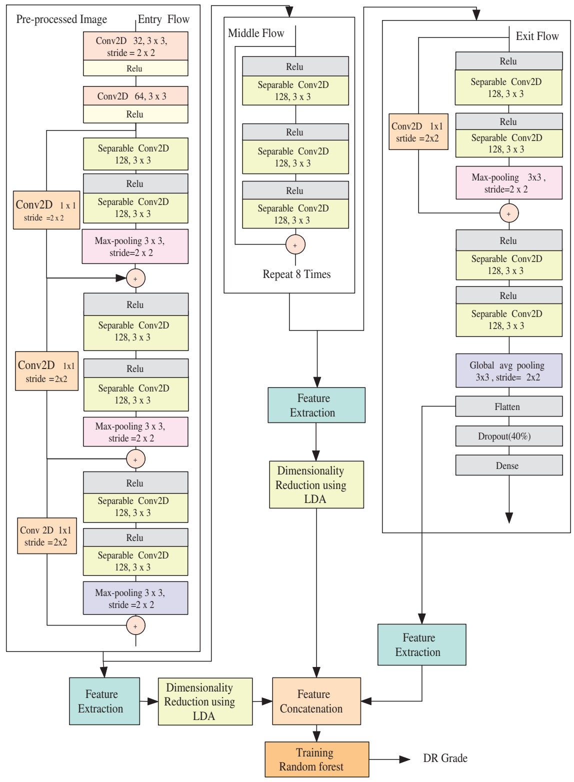

Fig. 2. Proposed network architecture.

tools and conditions. Image pre-processing is extremely crucial because image pre-processing quality impacts classification results. Several image processing techniques as sown in Fig. 1 are used to standardize these datasets in order to improve the model's performance. Cropping is done first to remove the extra black background in the fundus images. CLAHE [12] is a technique for equalizing images and limiting the amplification of unwanted noise. CLAHE is a variant of adaptive histogram equalization (AHE) and an image processing technique that increases image contrast . In order to redistribute the image's lightness values, AHE computes a number of histograms, each one corresponding to a different region of the image. Therefore, it is appropriate for enhancing the definition of edges in each area of an image as well as the local contrast. CLAHE, on the other hand, avoids over-amplification of contrast by operating on a small area of an image known as a tile rather than the entire image. By clipping the histogram at a predetermined value, CLAHE reduces amplification. The clip limit of a histogram is the value at which it is clipped, and any value that exceeds the clip limit is redistributed evenly among all of the histogram bins. CLAHE's parameters are clip limit = 2 and tile size = 4. to improve the accuracy, Pixel mean subtraction is carried out by calculating the mean image from the average of all train images' corresponding pixels and then subtracting the mean image from any other input image. Deep learning models require huge amounts of data to train them, but due to the limited number of images available, real-time data augmentation is used to train the models. Real-time data augmentation such as rotation, zooming, horizontal flip, and vertical flip is performed on train images so that the model receives a new image for each epoch during training.

C. *Arcitecture*

In the proposed framework, Xception pre-trained model [1] as the backbone is used for the extraction of deep features, which are then used to train a random forest for DR classification. The Xception model has three levels: entry flow, middle flow with eight repetition of same block, and exit flow as shown in Fig. 2 . The Xception network is modified by removing the top layer and replacing it with the dense layer, allowing the pre-trained weights from the ImageNet dataset to be transferred. The Xception network consists of depth-wise separable convolutions and linear residual connections. The Xception network is made up of 36 convolutional layers (including separable convolutional layers) divided into 14 modules. Batch normalization is performed after the convolutional and separable convolution operations in each module. With the exception of the first and last modules, 12 of the 14 modules are linked by a linear residual connection.

To improve classification accuracy, features are extracted (global and local features) from intermediate levels—after the entry-level, the middle level, and the flattened layer in the exit flow connection. Because there are more features in the entry and middle flows, the complexity of direct feature concatenation increases. As a result, LDA is used to reduce the number of features from the entry flow and middle flow in order to select the key characteristics for class separation and to reduce the complexity. LDA reduces the number of features by selecting the best features while retaining information for class separation. Following dimensionality reduction, the features extracted from various levels are concatenated and used to train a random forest for DR-grade classification.

1) depth-wise separable convolutions: depth-wise separable convolution layers in the Xception network divide the kernel into two distinct kernels, each of which performs two convolutions: depth-wise convolutions and point-wise convolutions. In contrast to conventional convolution, depth-wise convolution continues to perform convolution on each channel independently. In contrast, point-wise convolution uses a 1x1 kernel on all channels simultaneously. The separable convolutional operation reduces the overall computations and complexity of the network.

2) Dimensionality Reduction: LDA is a major strategy for dimensionality reduction techniques in machine learning. To achieve maximum class separability, LDA projects the data into a new feature space. LDA extracts p-independent features from data with d-independent features that best separate the classes. LDA generates fewer features than the original features, reducing computational cost and complexity. LDA generates two different scatter matrices: (1) a between-class matrix (S_b), and (2) within-class matrix (S_w). S_b determines the separation between the means of each category and S_w determines the separation between the means of each category as well as the data that represents that category [13].

$$s_b = \sum_{p=1}^{c} N_p(\mu_p - \mu)(\mu_p - \mu)^T \tag{1}$$

$$s_w = \sum_{p=1}^{c}\sum_{j=1}^{n} (x_j - \mu_p)(x_j - \mu_p)^T \tag{2}$$

In this case, c signifies the number of classes, n is the total number of samples, x_j is the j^{th} sample in a class c, μ indicates the overall mean, μ_p mean vector of the respective class, and N_p is the size of the respective classes.

In LDA, lower dimensional space is created by finding projection of K that maximizes (Sb) and minimizes (Sw).

$$\mathrm{Fisher's\ criterion} = \arg\max_{\mathrm{K}} \left| \frac{\mathrm{K^T S_b K}}{\mathrm{K^T S_w K}} \right| \tag{3}$$

where K is the projection matrix

3) Random Forest: Random Forest is a general class of ensemble methods built on decision trees. Decision trees are constructed utilizing a variety of tree predictors using a resampling method with replacement. It randomly samples the features and chooses the best split among those variables rather than choosing the best split among all features. An unknown instance's class is determined by majority vote. Many image classification researchers are interested in Random Forest because of its numerous advantages, including the ability to process a large number of input variables while remaining time-efficient.

III. Results

A. Evaluation metrics

Accuracy, precision, recall, F1-Score, and receiver operating characteristic (ROC) curve are the evaluation measures used to analyze how well the model performs. These metrics are numerically denoted by the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) [14]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\mathrm{F1-Score} = \frac{2 * \mathrm{Pr}\,ecsion * \mathrm{Recall}}{\mathrm{Pr}\,ecision + \mathrm{Recall}} \tag{7}$$

B. Results on APTOS-2019 dataset

The APTOS-2019 dataset is used to evaluate the proposed model. The dataset is divided into 80% training and 20% testing images, with 2929 training images and 733 test images divided into five classes. The experiment is carried out using the ADAM optimizer with a drop out of 50% and a learning rate of 0.0001. The Xception network is initially trained for 30 epochs using CLAHE pre-processing, after which the features are extracted from intermediate levels. LDA is used to reduce the dimension of global features to reduce the training time required to train random forest. The features are then concatenated and used to train a random forest.

Table III shows the model's performance in terms of precision, recall, and f1-score for multi-class classification. the overall weight average for score for precision , recall, and F1-score is 86.92 %, 87.17 %, and 87.03% respectively.

TABLE III
APTOS-2019 Dataset class-wise performance for multi-class classification

DR Stage	Precision	Recall	F1-score
No DR	96.23	97.54	96.88
Mild	77.46	79.71	78.57
Moderate	90.35	88.55	89.44
Severe	37.03	32.25	34.48
Proliferative	57.57	58.46	58.01
Weighted average	86.92	87.17	87.03

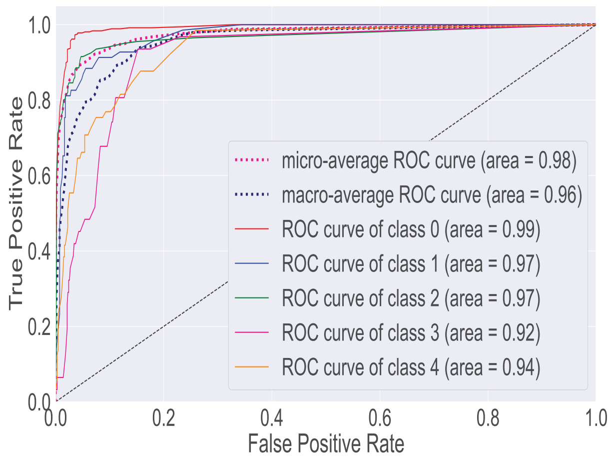

Fig. 3. ROC of the proposed technique.

ROC served as additional evidence of the suggested model's efficacy for multi-class classifications. The classifier uses the Area Under the Curve (AUC) score to differentiate among the

TABLE IV
PERFORMANCE COMPARISON

Literature	Accuracy (in %)	Precision (in %)	Recall (in %)	F1-score (in %)	AUC (in %)
Xception +SCL [1]	84.36	73.84	70.51	70.49	93.82
VGG16 + Xception + Gated-Attention [3]	82.54	82.00	83.00	82.00	79.00
DenseNet169+ CBAM [8]	82.00	-	-	68.00	-
Efficient MobileNetV2 + SVM [15]	85.00	-	-	-	93.00
Proposed Methodology	87.17	86.92	87.17	87.03	98.00

classes. Fig. 3 shows the AUC score for different class. the micro-averaged and macro-averaged AUC score is 98% and 96% respectively. The class wise AUC score for normal, mild, moderate, severe, and PDR is 99%, 97%, 97%, 92%, and 94% respectively. Fig. 4 shows the confusion matrix for the proposed model. Except for the severe class, the proposed model outperforms all others in classification.

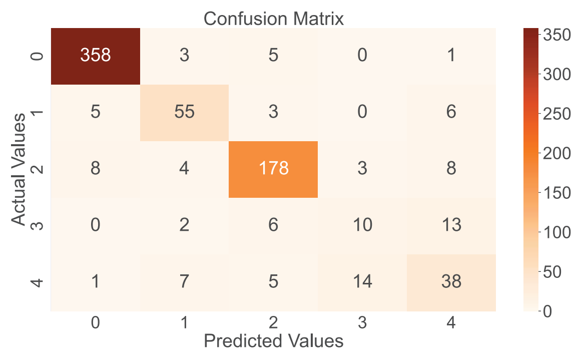

Fig. 4. Confusion matrix of the proposed method.

C. Comparison with Other Models

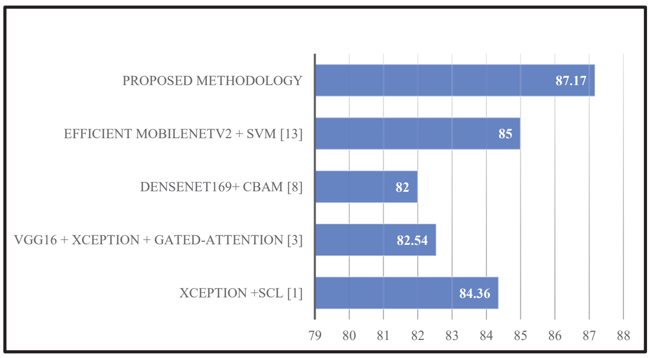

Fig. 5. Performance comparison of the proposed method with state-of-art techniques

Table IV compares the proposed model to current state-of-the-art models. In [1] supervised contrastive learning is used with Xception as the backbone and achieved 84.36% accuracy, 73.84% precision, 70.51% recall, 70.49% f1-score, and 93.82% AUC score, respectively. VGG16, and Xception models are utilised for the extraction of features, and these features are concatenated for further classification using gated- attention mechanism in [3]. DensNet169 is used as a backbone for CBAM and achieved accuracy and f1-score of 82% and 68% in [8]. How ever the authors did not mention precision, recall, and AUC score. In [15] extracted features from efficient MobileNetV2 are used to trained an SVM network for classification of DR, achieving 85% accuracy. In this work, the proposed methodology achieved an accuracy of 87.17%, precision of 86.92%, recall of 87.17%, and F1-score of 87.03%. These results demonstrate that the proposed method performs better in the classification of fundus images.

IV. CONCLUSION

In this work, a novel Xception based model with a combination of random forest classifier is proposed to classify DR. CLAHE-based pre-processing technique is used to improve the quality of the fundus images, as well as pixel-mean subtraction to improve the end-to-end performance of the proposed system. Xception architecture is utilized to extract global and local features from the fundus images. The dimensionality of these features is reduced by using LDA, which reduces the complexity and training time of the random forest classifier in the classification stage. The main advantages of the proposed method are (i) minimal training dataset (ii) less training time, and (iii) extraction of in-detail features from the fundus images. The proposed network is able to identify the class of the applied input with an accuracy of 87.17%, which is better compare to the existing techniques. The statistical metrics of the proposed method show that the performance is better compared to the recent state-of-the-art techniques.

REFERENCES

[1] M. R. Islam, L. F. Abdulrazak, M. Nahiduzzaman, M. O. F. Goni, M. S. Anower, M. Ahsan, J. Haider, and M. Kowalski, "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images," *Computers in Biology and Medicine*, p. 105602, 2022.

[2] C.-H. Hua, K. Kim, T. Huynh-The, J. I. You, S.-Y. Yu, T. Le-Tien, S.-H. Bae, and S. Lee, "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, 2020.

[3] J. D. Bodapati, N. S. Shaik, and V. Naralasetti, “Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 10, pp. 9825–9839, 2021.
[4] G. U. Nneji, J. Cai, J. Deng, H. N. Monday, M. A. Hossin, and S. Nahar, “Identification of diabetic retinopathy using weighted fusion deep learning based on dual-channel fundus scans,” *Diagnostics*, vol. 12, no. 2, p. 540, 2022.
[5] K. Shankar, Y. Zhang, Y. Liu, L. Wu, and C.-H. Chen, “Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification,” *IEEE Access*, vol. 8, pp. 118 164–118 173, 2020.
[6] P. Saranya and S. Prabakaran, “Automatic detection of non-proliferative diabetic retinopathy in retinal fundus images using convolution neural network,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–10, 2020.
[7] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, “Automatic detection of diabetic eye disease through deep learning using fundus images: a survey,” *IEEE Access*, vol. 8, pp. 151 133–151 149, 2020.
[8] M. M. Farag, M. Fouad, and A. T. Abdel-Hamid, “Automatic severity classification of diabetic retinopathy based on densenet and convolutional block attention module,” *IEEE Access*, vol. 10, pp. 38 299–38 308, 2022.
[9] H. Kaushik, D. Singh, M. Kaur, H. Alshazly, A. Zaguia, and H. Hamam, “Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models,” *IEEE Access*, vol. 9, pp. 108 276–108 292, 2021.
[10] S. Gayathri, V. P. Gopi, and P. Palanisamy, “Automated classification of diabetic retinopathy through reliable feature selection,” *Physical and Engineering Sciences in Medicine*, vol. 43, no. 3, pp. 927–945, 2020.
[11] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. Prasad, “Diabetic retinopathy detection and classification using capsule networks,” *Complex & Intelligent Systems*, pp. 1–14, 2021.
[12] S. H. Rasta, M. E. Partovi, H. Seyedarabi, and A. Javadzadeh, “A comparative study on preprocessing techniques in diabetic retinopathy retinal images: illumination correction and contrast enhancement,” *Journal of Medical signals and sensors*, vol. 5, no. 1, p. 40, 2015.
[13] R. F. Mansour, “Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy,” *Biomedical engineering letters*, vol. 8, no. 1, pp. 41–57, 2018.
[14] J. P. Allam, S. Samantray, and S. Ari, “Spec: A system for patient specific ecg beat classification using deep residual network,” *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1446–1457, 2020.
[15] S. c, A. Handayani, B. R. Hermanto, and T. L. E. R. Mengko, “Diabetic retinopathy classification using a hybrid and efficient mobilenetv2-svm model,” in *2020 IEEE REGION 10 CONFERENCE (TENCON)*. IEEE, 2020, pp. 235–240.