IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 September 2023, accepted 25 October 2023, date of publication 2 November 2023, date of current version 9 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3330104

RESEARCH ARTICLE

A Faster RCNN-Based Diabetic Retinopathy Detection Method Using Fused Features From Retina Images

MD. NUR-A-ALAM[1,3], MD. MOSTOFA KAMAL NASIR[1], MOMINUL AHSAN[2], MD. ABDUL BASED[3], (Member, IEEE), JULFIKAR HAIDER[4], AND SIVAPRAKASAM PALANI[5]

[1]Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Tangail 1902, Bangladesh
[2]Department of Computer Science, University of York, YO10 5GH York, U.K.
[3]Department of Electrical, Electronics and Telecommunication Engineering, Dhaka International University, Dhaka 1205, Bangladesh
[4]Department of Engineering, Manchester Metropolitan University, M1 5GD Manchester, U.K.
[5]College of Electrical and Mechanical Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

Corresponding author: Sivaprakasam Palani (shiva@aastu.edu.et)

ABSTRACT Early identification of diabetic retinopathy (DR) is critical as it shows few symptoms at the primary stages due to the nature of its gradual and slow growth. DR must be detected at the early stage to receive appropriate treatment, which can prevent the condition from escalating to severe vision loss problems. The current study proposes an automatic and intelligent system to classify DR or normal condition from retina fundus images (FI). Firstly, the relevant FIs were pre-processed, followed by extracting discriminating features using histograms of oriented gradient (HOG), Shearlet transform, and Region-Based Convolutional Neural Network (RCNN) from FIs and merging them as one fused feature vector. By using the fused features, a machine learning (ML) based faster RCNN classifier was employed to identify the DR condition and DR lesions. An extended experiment was carried out by employing binary classification (normal and DR) from three publicly available datasets. With a testing accuracy of 98.58%, specificity of 97.12%, and sensitivity of 95.72%, this proposed faster RCNN deep learning technique with feature fusion ensured a satisfactory performance in identifying the DR compared to the relevant state-of-the-art works. By using a generalization validation strategy, this fusion-based method achieved a competitive performance with a detection accuracy of 95.75%.

INDEX TERMS Diabetic retinopathy (DR), fundus image, histograms of oriented gradient (HOG), Shear let transform, faster RCNN, feature fusion.

I. INTRODUCTION

Diabetes mellitus is a category of metabolic disorders marked by chronic high blood sugar levels [1]. This raises blood glucose levels to a dangerously high level, and therefore, the body seems unable to create enough insulin under this circumstance or the body cannot respond to the insulin that is produced [2]. Many life-threatening diseases such as blindness, kidney failure, heart attacks, stroke and lower limb amputation could be caused by diabetes. According to the latest IDF statistic, about 547 million individuals will be affected globally by diabetes in 2022, accounting for almost 10% of all adults aged between 20 to 79. In 2030, this number is forecast to rise to reach 634 million and by 2045 to reach 786 million. Furthermore, it is estimated that 240 million people with diabetes are now undiagnosed [3]. Likewise, caused by diabetes, approximately 1.6 million people die immediately [4]. In developing countries like Bangladesh, around 13.1 million people are affected by diabetes, due to lack of awareness and proper diagnosis [2]. The quantity of diabetes patients is rising every day, and so the mortality is.

Diabetic Retinopathy (DR) is a diabetic condition caused by hyperglycemia, or high blood sugar level [5]. DR causes the retina's blood vessels to enlarge and leak blood and other

The associate editor coordinating the review of this manuscript and approving it for publication was Gyorgy Eigner.

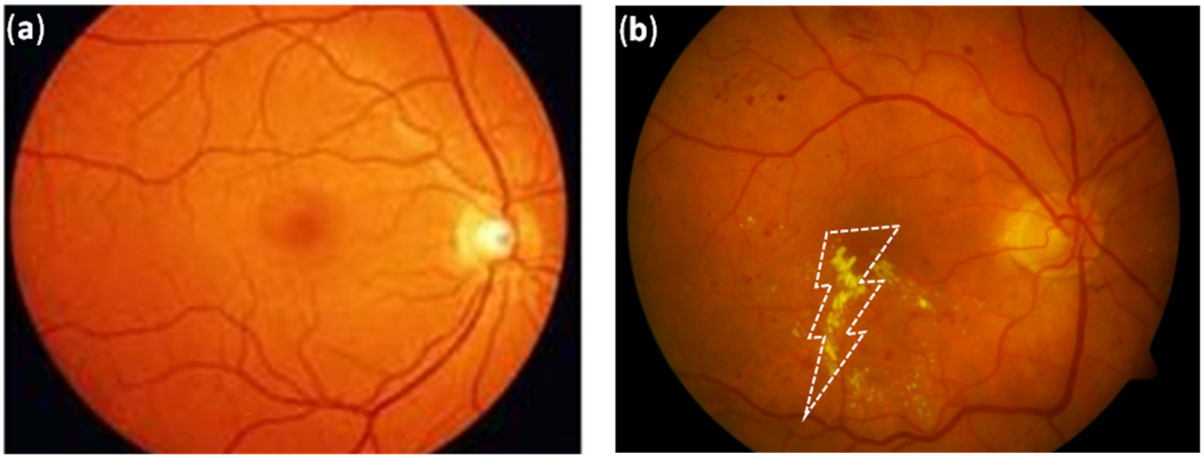


FIGURE 1. Comparison between (a) normal and (b) Diabetic Retinopathy (DR) fundus images.

TABLE 1. List of existing works for diabetic detection.

Ref.	Methods	Dataset (images)	Finding (Accuracy)
[17]	Faster R-CNN	Kaggle (4,000)	OD Localization and Segmentation (85.4%)
[20]	Discrete wavelength + texture + Statistical feature + Random Forest	Self-prepared (338)	Classification (89.63%)
[12]	Gaussian Mixture Model (GMM)+KNN+ SVM	MESSIDOR (1,200)	Lesion Classification (92.34%)
[19]	DenseNet-65+ Faster-RCNN	Kaggle (88,704)	Lesion localization + Classification (97.2%)
[22]	Hybrid Faster R-CNN+MD-ResNet	HEI-MED (3,000)	Grading Diabetic macular edema (DME) (94.4%)
[23]	Enhanced Multi-feature Fusion Network (EMFN)	MESSIDOR (4,500)	Detection Hard Exudate (94%)
[29]	HPTI-v4	MESSIDOR (1,200)	DR detection (99.49%)
[30]	PCA + Firefly algorithm + DNN models	UCI repository (200)	DR detection (97%)
[44]	local cross-section transformation (LCT)+boosting-based classifier (RUSBoost)	DiaretDB1(89) + e-ophtha-MA (148) + ROC(100)	Retinal microaneurysm (MA) detection (93.67%)
[43]	Deeply supervised ResNet + modified ResNet architecture	Kaggle (5,000)	DR grading (81%)
[45]	VGG16-fc1, Xception network and CGG16-fc2 base deep feature + DNN	DDR (380)	DR identification (80.96%)
[33]	GoogleNet, AlexNet, ResNet, and VGG16	Kaggle (5,000)	Classification (95.68%)

body fluids [6]. If DR is left undiagnosed and untreated, this might result in significant eyesight loss. DR usually takes many years to progress to the point where it can cause total blindness. DR is involved in 2.6% of blindness worldwide [7]. Patients who have been suffering from uncontrolled diabetes for a long time are more likely to have the DR. Diabetics should have regular retina screenings to identify and treat DR early enough to avert blindness [8]. The presence of various lesions on a retina picture, for instance, Micro aneurysms (MA), internal bleeding, and hard and soft exudates (EX), is used to detect the DR [9], [10]. The comparison of both normal and DR retina images is depicted in Figure 1. In normal retina images, the components of the retina (macula, fovea, and blood vessels) are visible, but the image in the diabetic retina is physical distortion by deep vein thrombosis (blood clots) and yellow pigmentation [11]. DR must be identified and classified in a timely and precise manner. Because of low-intensity information in the retina fundus images (FI), detecting the DR lesions remains a challenging task.

A large number of research has been conducted on DR detection using different deep learning approaches with features extracted and classified with a variety of techniques

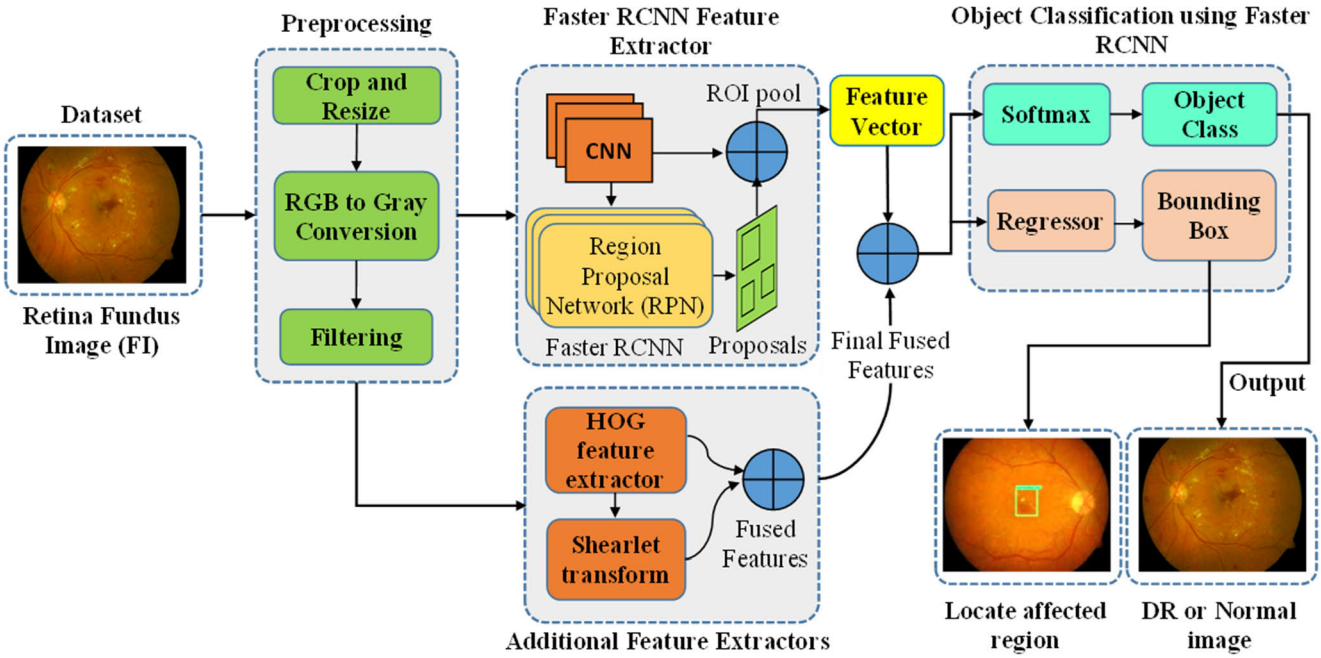


FIGURE 2. Illustration of the proposed Faster-RCNN architecture for DR identification.

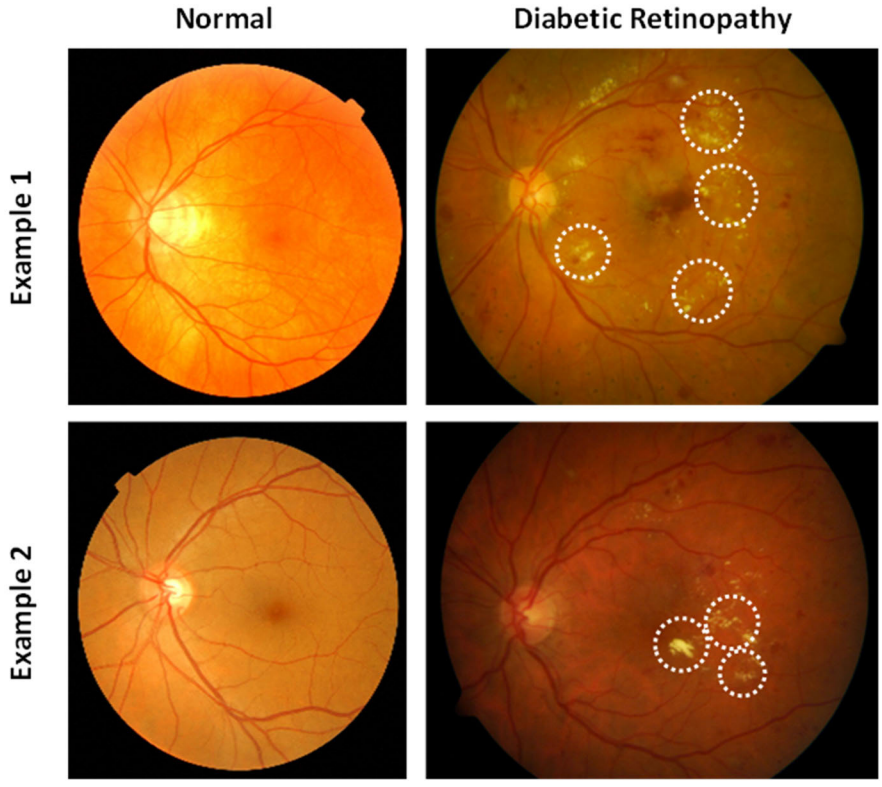


FIGURE 3. Dataset image samples of (a) normal (b) diabetic retinopathy (DR) images.

[7], [18], [19], [20]. Zago et al. [21] presented a CNN-based solution for automated detection of DR using the CNN approach, which pinpointed the location of the wound lesions in the FIs. The developed model produced good results (ACC=95.78%).

DR detection from retinal pictures based on CNN, and the random forest was proposed by Wu et al. [22]. The severity of the DR was identified using the presented approach with the Messidor-2 dataset. In another study, a synergic deep learning approach-based automated technique for DR detection and

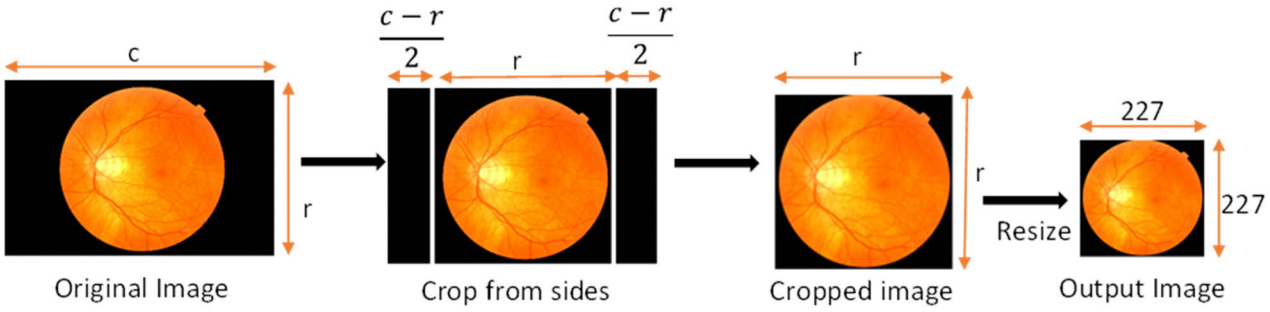


FIGURE 4. Cropped and resized FI image during the pre-processing stage.

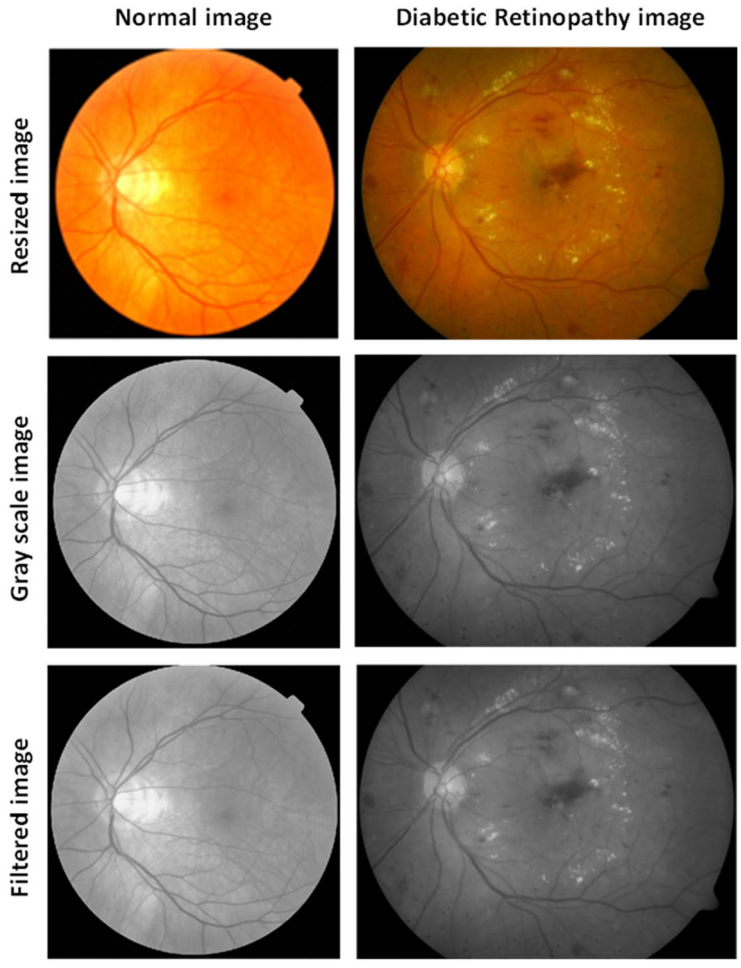


FIGURE 5. Preprocessing of the input FIs.

classification was presented by Guo et al. [23]. DenseNet, ResNet50, and VGG16 models were used for detecting DR lesions [24]. The approach was computationally sound. However, it may not accurately categorize microaneurysms due to the presence of fluorescein, which makes them readily misclassifiable.

Leeza et al. [18] proposed a method for detecting DR lesions automatically by incorporating the pathological qualities of the input image into a bag of features (BOF) concept. SURF and HOG descriptors were utilized to compute the critical points of an input sample, which were then merged to form a dictionary. Furthermore, the generated dictionary was subjected to coding and pooling to produce a compressed representation of the characteristics.

An approach to automatically classify DR lesions was given by Seoud et al. [13] using the idea of dynamic form

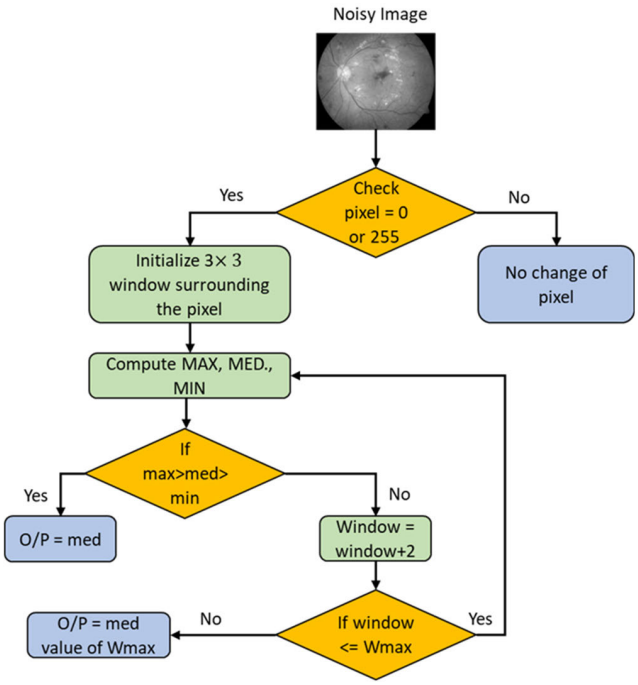


FIGURE 6. Flowchart of adaptive mean filtering of noisy DR image.

characteristics (DSF). The features are extracted using a multi-scale ring-shaped matching filter. An RF classifier is used to perform the classification. The main limitation of this research is only red lesions can be found with this method. A method for DME identification based on the Gaussian and difference of Gaussian filter bank was introduced by Marin et al. [14]. The potential DME regions are classified using a regularized local regression (RLR) classifier. A DL technique was presented by Abramoff et al. [15] for the automatic identification of DR. Comparing the approach in [15] to straightforward hand-coded-based solutions, a superior performance was reported. The DL technique was used by Gargeya et al. [16] to develop a completely automated system for DR recognition. Finally, CNN and SVM were used to classify the data. This study was effective in DR detection; nevertheless, many samples were misclassified because they did not account for the DR with maculopathy. Table 1 summarizes other relevant research reported in the literature on the DR detection. Although the evidence of successful DR classification has been presented with a variety of DL models, the feature fusion technique has not been studied extensively.

In this research, a machine learning method has been proposed based on Faster-RCNN to concurrently detect DR anomalies by calculating the deep features of the input FIs and locating the impacted region for overcoming the limitations of the existing methods. The working procedure of this proposed system is preprocessing, filtering, feature extraction, fused features, classification, and localization. This work's main novelty lies in creating a fused feature vector by combining features extracted by three techniques to achieve better DR detection performance by a Faster RCNN classifier. The significant novelty of this research can be outlined below.

(1) A new dataset of 21,171 FIs was developed using three publicly available datasets.

(2) An adaptive filter was used to eliminate noise from the retina images, which can assist in obtaining better classification output.

(3) HOG, Shearlet transform and RCNN were used to extract discriminating features and to create a fused feature vector.

(4) A Faster RCNN-based feature fusion model was developed to detect DR anomalies and locations of the affected DR lesions.

(5) The results were compared to the state-of-the-art methods for DR classification available in the literature.

The remaining sections of the paper are organized as follows. Section II discusses the proposed deep learning architecture for the DR classification. Section III provides details on the experiments conducted, results obtained and analysis of the outcomes. Finally, Section IV draws conclusions based on the findings.

II. PROPOSED DIABETICS CLASSIFICATION METHODOLOGY

Detecting DR from FI is a four-way process: preprocessing, feature extraction, classification, and localization between the DR anomalies and healthy images. In the first stage, the images for the dataset were collected, appropriately categorized, and merged into a single dataset. It was preprocessing involved cropping unwanted regions from the original image, resizing to a standard size, and transforming the original data set's RGB images to grayscale images. After performing annotations, the Faster-RCNN approach [26] was employed to extract characteristic features from the FIs. In addition, features were also extracted individually by Histogram of oriented gradient (HOG) and Shearlet transform, and all three feature vectors were fused. Finally, a Faster-RCNN was used to classify the DR and normal images and locate the DR anomalies within the images. The architecture of the suggested approach is depicted in Figure 2.

A. DATASET DEVELOPMENT

A number of datasets are publicly available for detecting DR and normal images in the retina. They are frequently utilized for training, to test and validate different ML algorithms employed, and to compare the performance of one system to that of the others. As a benchmark for evaluating the intelligent system's performance, the datasets were classified as DR or normal retina images. In this work, three standard datasets were used for DR detection.

1) DIARETDB1 [24]

The dataset includes sample retina FIs with a resolution of 1500×1152 pixels and a field of vision (FOV) of 50 degrees for 89 distinct candidates. Professionally anointed DR images (84) and normal images (5) are included.

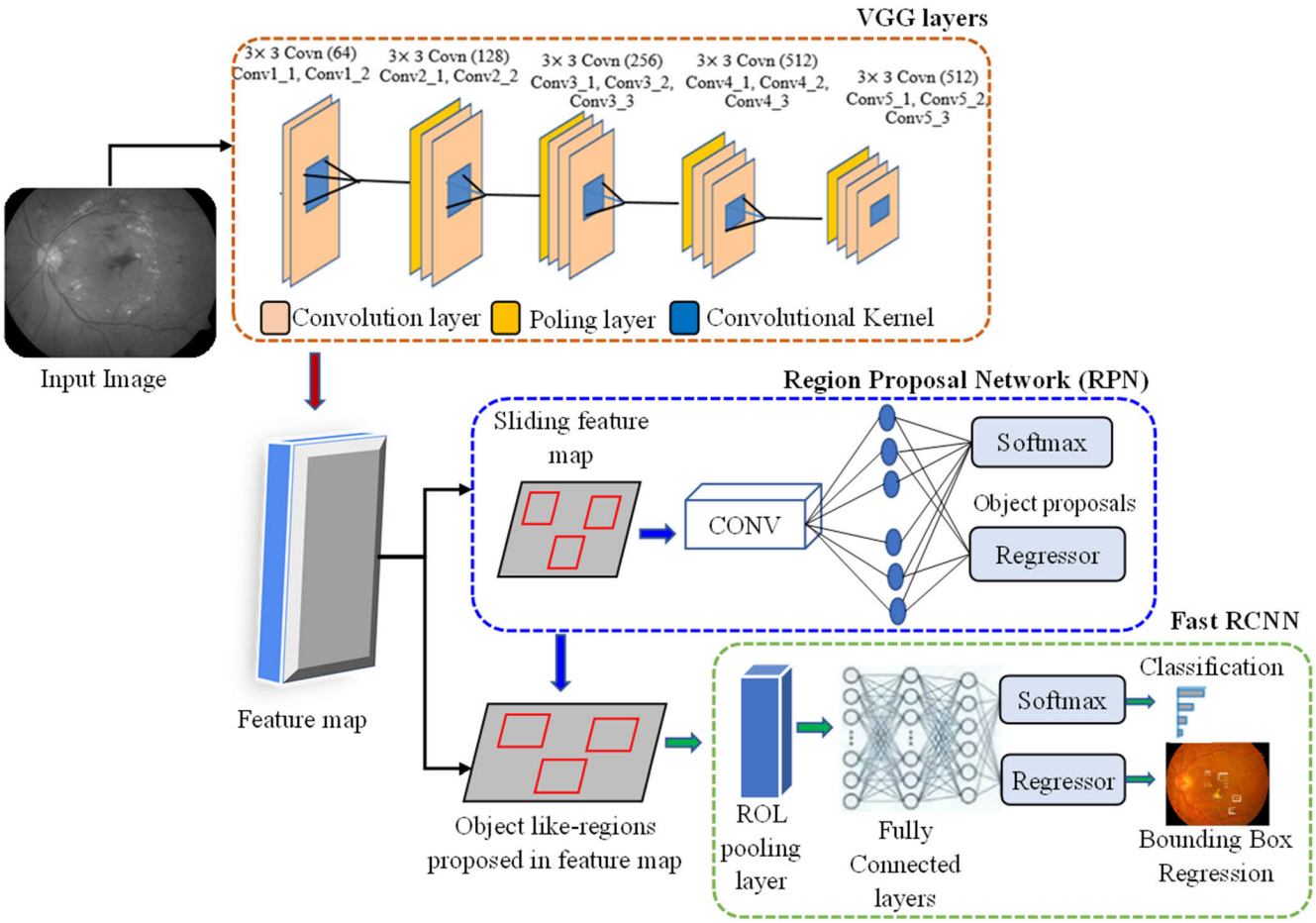


FIGURE 7. Illustration of the Faster RCNN architecture.

2) KAGGLE [25]

The dataset contains 20,702 high-resolution retina images captured from several cameras with dimensions ranging from 433 × 289 pixels to 1184 × 1456 pixels. In this dataset, 10,500 images represent the DR, and the rest of the 10,202 images are normal images. The ground facts for training images are provided to the public. Many of the images on Kaggle are of poor quality and have inaccurate tagging [22], [26].

3) DDR [22]

These publicly available images were used to segment blood vessels. It provides 45 images with a size range of 1504 × 1336 pixels. There are 200 images with DR, 180 images with healthy eyes, and 15 images with glaucoma.

The three datasets were merged in one single dataset containing a total of 21,171 fundus retina images (excluding 15 glaucoma images) and divided into 10,784 images as the DR images and 10,387 images as the normal images. All images were in RGB format and different sizes ranging from 433 × 289 to 1504 × 1456 pixels. Figure 3 shows sample images of DR and normal categories. The highlighting regions represent that the FI with diabetes is diagnosed where the image is fractured.

B. PREPROCESSING

One of the most important aspects of medical image analysis is data preparation, which aids in extracting valuable characteristics. The image size was transformed to 227 × 227×3 pixels and converted to grayscale (227 × 227×1) pixels from the RGB format. An adaptive median filtering was utilized for denoising and making the scaled image uniform. In this step, each original image was first squared by cropping from the left and right sides. The cropped square images were then down sampled to 227 × 227×3, as shown in Figure 4. Finally, the RGB images were converted to the grayscale format, as displayed in Figure 5.

The existence of superfluous information in the training images can influence the proposed network's detection performance. The input images were subjected to a light correction process utilizing an adaptive median filtering (AMF) approach [24], which eliminated noise, improved contrast, and restored the image quality (Figure 5). This study used a new technique for designing the AMF to remove impulsive

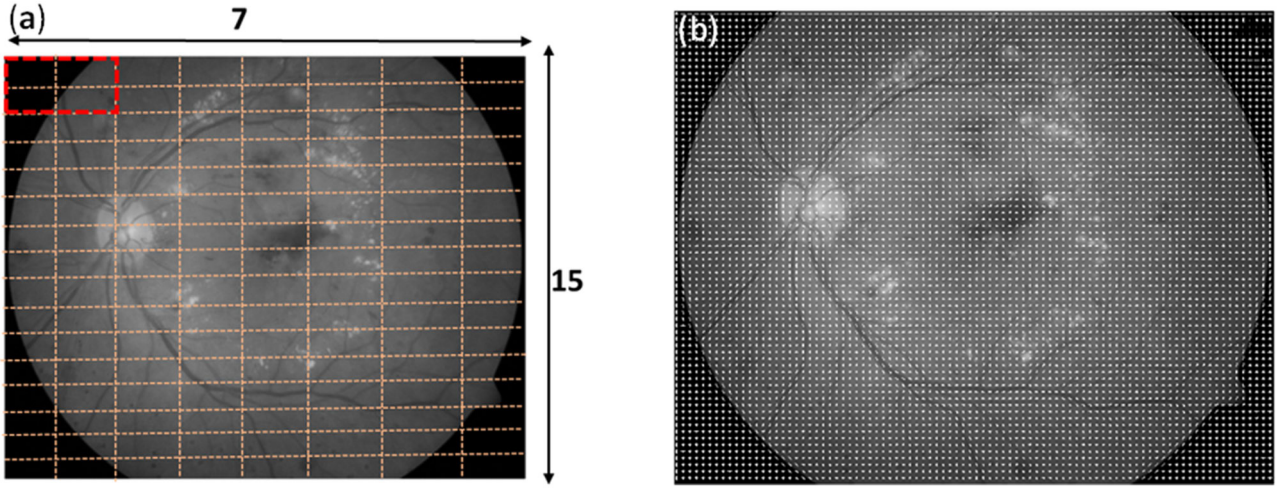


FIGURE 8. Calculated (a) HOG features and (b) histogram of oriented gradients.

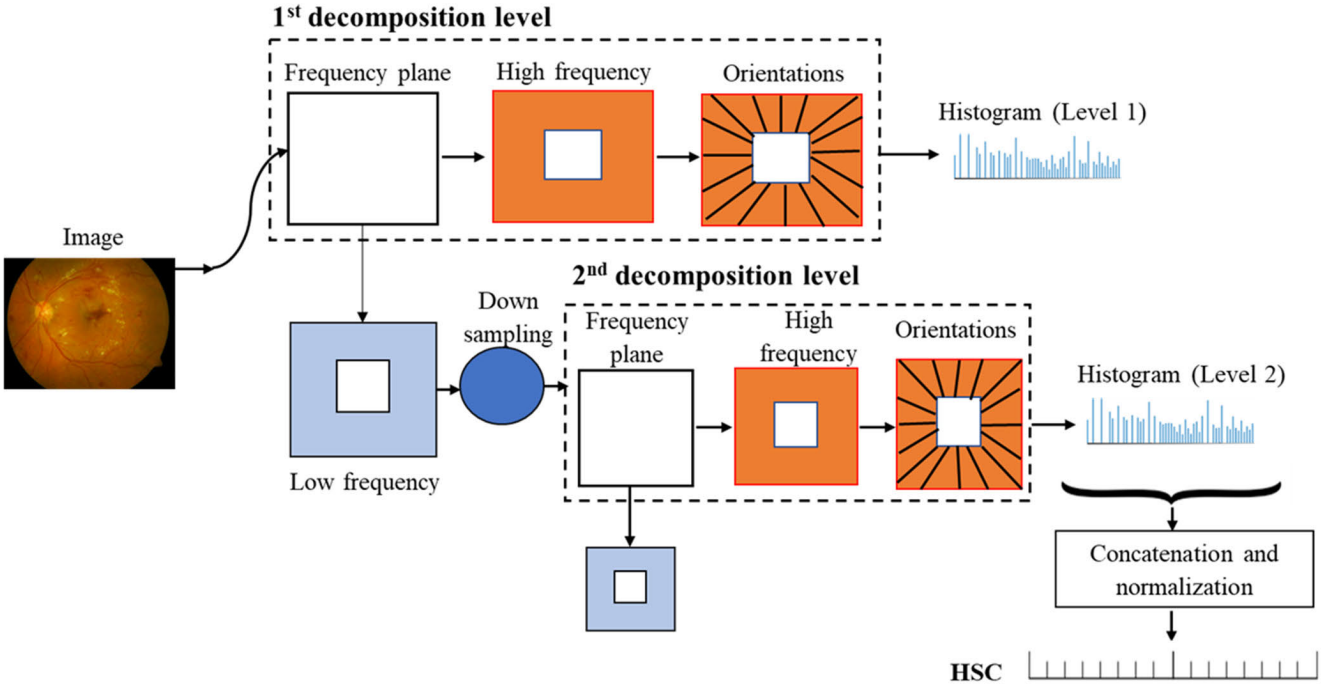
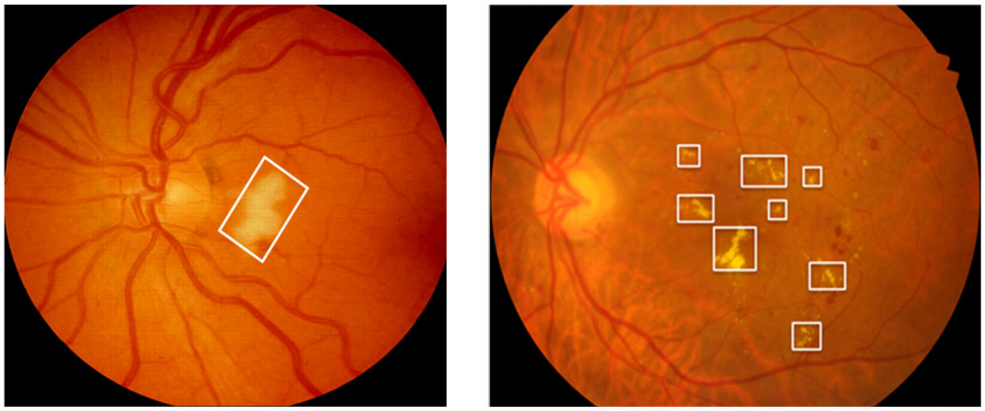


FIGURE 9. Feature extraction procedure by Shearlet transform (ST).

noise and minimize image distortion. The noise reduction filter was meant to reorganize itself in real-time. Therefore, a high-speed non-linear adaptive median filter was implemented in this study. The filtering process can improve an image damaged with impulse noise by up to 70% using the AMF. The images were filtered by following the AMF algorithm presented in Figure 6.

The adaptive filter operates in two steps, first determining the kernel's median value and then if the current pixel value is an impulse or not. If a pixel's value is corrupted, it either modifies it using the median or keeps the grayscale pixel value.

The maximum window size has been initialized, Wmax = 39 × 39 (original image). If any pixel was found to be a noisy one then initialized the current window size 3 × 3 (mask) surrounding the noisy pixel and computed the minimum, median and maximum value of the window. If maximum>median>minimum, then the output pixel is equal to the median value of the window. Otherwise, the window size is increased by 2 and repeat the process with



(a) DR cotton wool retina image (b) DR hard exudates retina image

FIGURE 10. Detected cotton wool and hard exudates spots in the DR images.

TABLE 2. Distribution of dataset for training, testing and validation.

Datasets	DR	Normal
Training	8,628	8,310
Testing	2,156	2,077
Total	10,784	10,387

increase of window size to get a suitable output pixel and this process is continued until the window size is less than equal to the maximum window. If the window size is more than the maximum window then the output pixel is equal to the median value of the maximum window.

C. FASTER RCNN FEATURE EXTRACTOR

The faster RCNN classifier consists of two modules: a deep, fully connected network, a region proposal network (RPN), and a fast RCNN detector. The region proposal network's mission is to support the discovery of the best areas of various sizes and proportions. To build the collections of regions, faster R-CNN employs a region proposal approach. Faster R-CNN has an additional CNN for obtaining regional proposals, which is referred to as the regional proposal network. The proposal network in the training region accepts the feature map as input and produces region proposals. These recommendations are then forwarded to the ROI pooling layer for further processing [43]. The architecture of the faster RCNN is shown in Figure 7. The four steps of the faster RCNN approach are convolution layer, RPN, ROI pooling, and classification.

The Faster RCNN comprises 13 convolutional and ReLu layers and four pooling layers. The Faster RCNN network uses these convolutional layers to produce the feature map of the input picture, which is subsequently communicated with the RPN module and other layers. The RPN module is made up of 3×3 fully convolutional layers that are utilized to construct anchors and bounding box regression offsets. The Softmax function is adopted to determine whether the computed anchors lie in the background or foreground. The object suggestions are computed using the produced anchors and bounding boxes at the final stage. This layer generates proposal feature maps and shares them with all the related network layers. Finally, the categorization stage determines the spots of diabetics found. It operates by utilizing the ROI pooling layer's output. The resultant position of the identified test box is displayed via bounding box regression.

D. HISTOGRAM OF ORIENTED GRADIENTS (HOG) FEATURE EXTRACTOR

The suggested approach employed a greater number of histogram bins on distinct parts of the pictures to retrieve HOG features [42]. The gradient of an image can be expressed as ∇f for an image f(x, y) in Equation 1.

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{1}$$

Here, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ is the image's derivative with respect to x and y. Equations 2 and 3 can be used to determine derivatives.

$$f_x(x) = \frac{\partial f}{\partial x} = f(x, +, 1) - f(x, -, 1), \tag{2}$$

$$f_y(y) = \frac{\partial f}{\partial y} = f(y, +, 1) - f(y, -, 1). \tag{3}$$

The derivatives are calculated in practice by convoluting the pictures along the x and y axes with the kernels [−1 0 1] and $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. The magnitude (g) and direction (θ) of the gradients can be estimated using Equation 4 and 5 after the gradients have been calculated.

$$g = \sqrt{g_x^2 + g_y^2} \tag{4}$$

$$\theta = tan^{-1} \frac{g_y}{g_x} \tag{5}$$

TABLE 3. Comparison study of different filtering methods.

Performance Metrics for all filter techniques	Average filter	Median filter	Mean Filter	Adaptive median filter
MSE	227.5623	0.8563	230.8934	0.5672
PSNR	25.4312	38.7232	18.5434	48.7342
NK	0.0756	0.0018	0.0893	9.2332e-4

TABLE 4. Experiment results of performance based on different fused methods with Faster RCNN classification.

Individual and fused features	ACC	SP	SE	PR
Using SIFT Features	0.7954	0.7834	0.7712	0.7456
Using HOG Features	0.8654	0.8765	0.8867	0.8754
Shear let Transform Features	0.8356	0.8234	0.8345	0.8378
Faster RCNN features only	0.9656	0.9389	0.9234	0.9457
(HOG + Faster RCNN) features	0.9778	0.9356	0.9478	0.9321
(Shear let Transform + Faster RCNN) features	0.9534	0.9122	0.9032	0.9172
(HOG+ Shear let Transform+ Faster RCNN) features	0.9858	0.9857	0.9861	0.9859

TABLE 5. Comparison outcomes for several CNN based feature extractors with Faster RCNN classification.

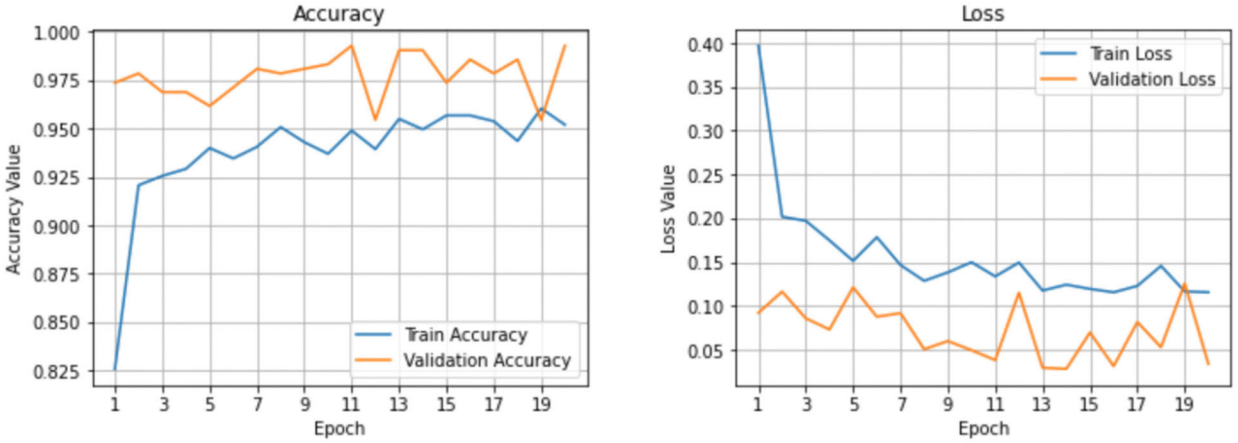
Fused features with CNN based extractor models	ACC (%)	SP (%)	SE (%)	PR (%)
HOG+ Shear let Transform +LSTM	92.79	91.17	91.53	90.85
HOG+ Shear let Transform +VGG16	93.90	92.50	94.17	91.59
HOG+ Shear let Transform +ResNet50	95.43	92.31	94.56	93.37
HOG+ Shear let Transform +VGG19	96.67	93.23	95.45	96.72
HOG+ Shear let Transform +RCNN	94.89	93.78	92.45	94.67
HOG+ Shear let Transform +Fast RCNN	97.89	96.43	94.78	95.23
HOG+ Shear let Transform +Faster RCNN	98.58	98.57	98.61	98.59

TABLE 6. Comparison performance for classification using different CNN models with features extracted by HOG+ Shear let Transform +Faster RCNN.

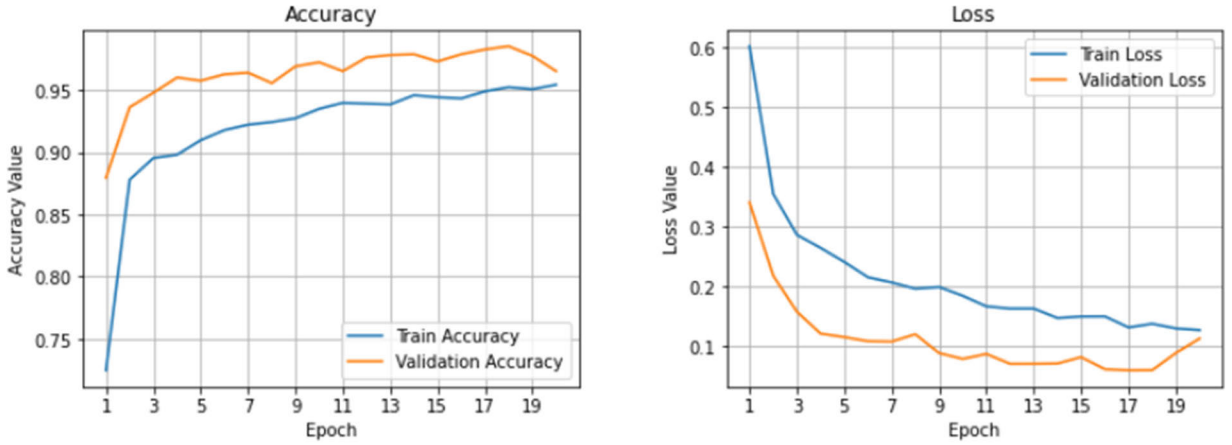
Classifiers	ACC (%)	SP (%)	SE (%)	PR (%)
VGG19	96.67	94.27	95.49	95.10
RCNN	94.89	94.28	95.13	94.33
Fast RCNN	97.19	96.64	97.71	96.77
Faster RCNN	98.58	98.57	98.61	98.59

Blobs from retina images are clipped and shrunk to $(64 \times 128 \times 1)$ before HOG is calculated. Figure 8(b) illustrates the generated histogram of directed gradients for a retina image from the dataset.

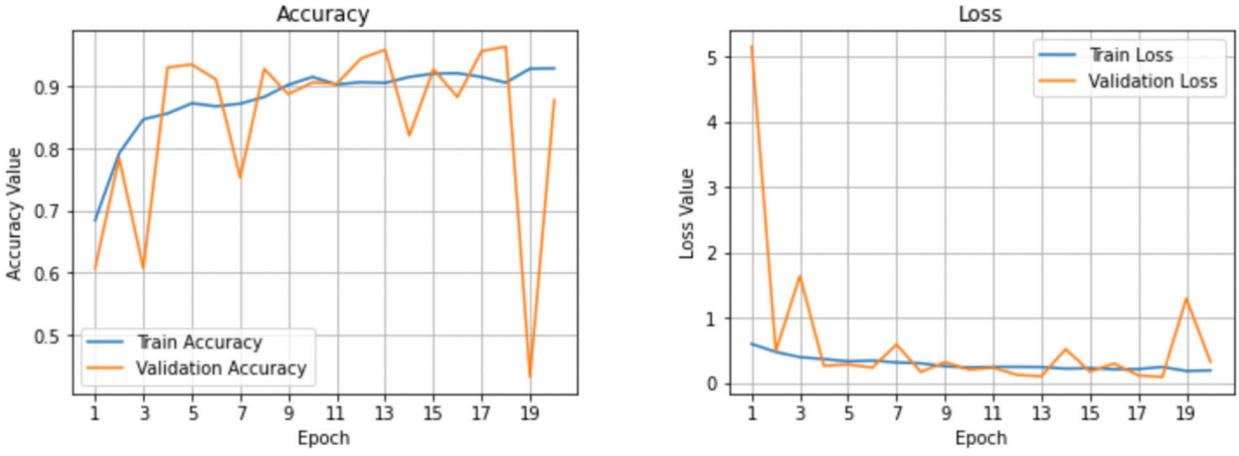
The scaled blobs are then separated into (8×8) cells, with each cell's gradient histogram is calculated along the x and y axes. By doing so, the features (or histogram) are obtained for the smaller patches which in turn represent the whole image. This value can be certainly changed from 8×8 to 16×16 or 32×32. If the image is divided into 8×8 cells and generate the histograms, a 9×1 matrix will be obtained for each cell. The frequency of certain values is shown via histograms. The histogram is calculated for nine angles with 20 degree apart: 0, 20, 40, 60, 80, 100, 120, 140, and 160 degrees. The



(a)



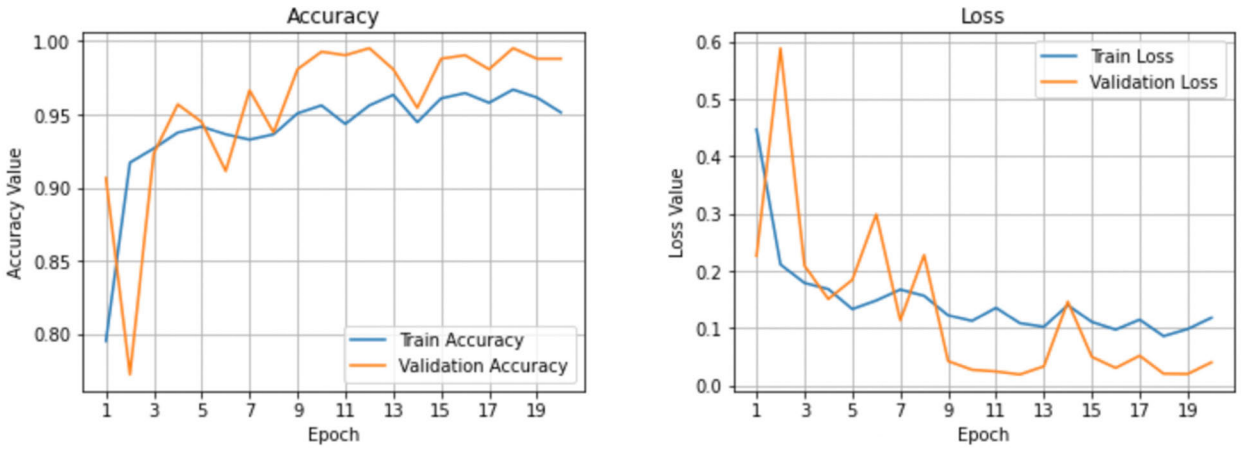
(b)



(c)

FIGURE 11. Acuracy and loss curves of (a) VGG19 (b) RCNN (c) fast RCNN.

value of the gradient magnitude is multiplied by the value of the matching gradient direction value in the histogram. The histogram output is represented as a 1×9 dimensional array. As a result, the histograms are normalized after calculation



(d)

FIGURE 11. *(Continued.)* **Acuracy and loss curves of (d) faster RCNN models during training and validation.**

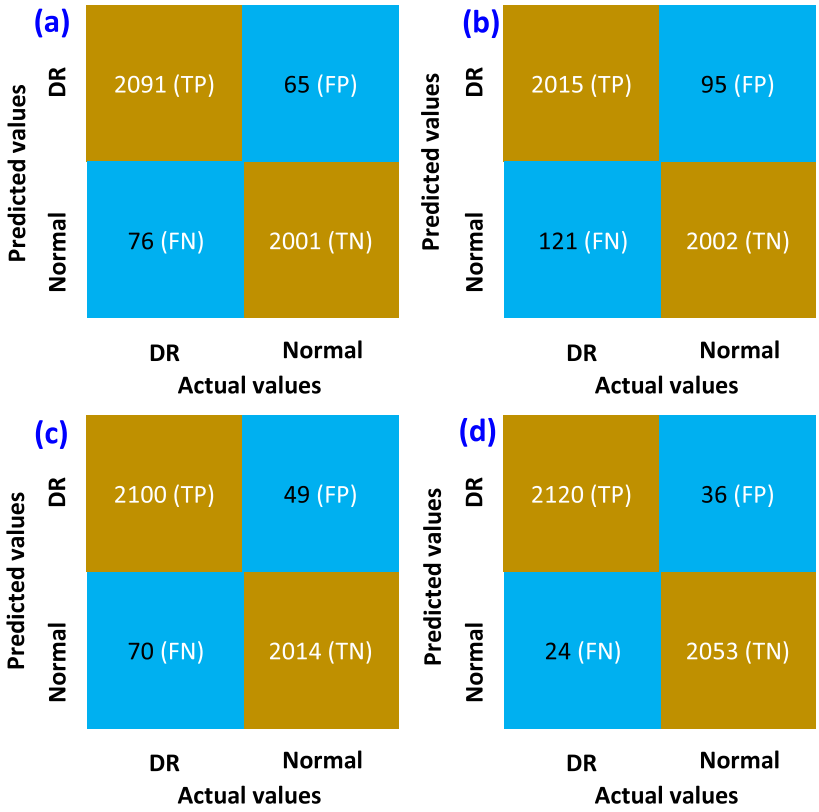


FIGURE 12. **Confusion matrixes of the overall framework with fused feature for (a) VGG19 (b) RCNN (c) fast RCNN (d) faster RCNN models.**

to lessen the system's susceptibility to overall lighting. L2 normalization is typically performed on (16 × 16) blocks. Four (8 × 8) gradient histograms are included in a (16 × 16) block with no overlaps. The histograms are concatenated to form a (1 × 36) element array. The histogram H are concatenated to form a (1 × 36) element array. And normalized the histogram H_norm. The L2 norm of a vector H = [h1, h2, h3, h4,.....h_n] can be calculated by Equation 6 and 7.

$$H_{l2} = |H| = \sqrt{h_1^2 + h_2^2 + h_3^2 + h_4^2} \tag{6}$$

$$H_{norm} = \frac{H}{|H|} \tag{7}$$

These normalized vectors' characteristics are then employed to carry out different computer vision tasks.

TABLE 7. Measure the several RCNN techniques.

RCNN Techniques	mAP	IoU	Time (s)
RCNN	0.897	0.918	0.27
Fast RCNN	0.945	0.945	0.24
Faster RCNN	0.965	0.975	0.20

In Figure 8(a), there are either (7 × 15) or 105 blocks of (16 × 16). A vector of size (1 × 36) is contained in each of these 105 blocks. As a result, the total features in an image would be 105 × 36=3,780. The features are subsequently sent to a CNN for object identification. Finally, a total of 3,780 accurate features were obtained from HOG.

E. SHEARLET TRANSFORM FEATURE EXTRACTOR

The Shearlet transform (ST) is a common framework for analyzing and displaying anisotropic information in an input image at many scales [22]. This method is useful for detecting an edge in the images and identifying their locations. The non-subsampled ST is illustrated in Figure 9. The extraction of features for a two-level Shearlet decomposition with eight orientations was conducted. For each level of decomposition, a histogram with the number of the same bins and orientations is computed by applying the ST. These histograms represent the distribution of edge orientations at the specified scales. From Shearlet coefficients at the relevant orientation, the absolute value is determined and combined to get each histogram bin's entry. Concatenation of the histograms computed for each decomposition level, followed by L2-norm normalization, yields the histograms of shearlet coefficients (HSC). Finally, thirteen features such as energy, mean, median, entropy, homogeneity, correlation, and others were obtained from the value of the shearlet coefficient.

F. FUSED FEATURES

A fusion-feature vector was created by combining the extracted features from HOG, Shearlet transform (ST) and RCNN. As previously stated, overlapping, redundancy, and dimensional increase are regular occurrences in all fusion-based techniques. The final fused features were used to classify the DR from the retina FIs. HOG, ST and RCNN retrieved features in Equations (8)-(10), as shown at the bottom of the next page, respectively. Concatenation is used to combine the retrieved feature vectors, represented by Equation (11), as shown at the bottom of the next page.

G. FASTER RCNN CLASSIFICATION AND LOCATING AFFECTED REGION

Finally, the classification stage was used to determine the DR or normal image by utilizing the ROI pooling layer's output. The faster RCNN classifier method classify the DR or normal images. After classifying the DR retina images the system locating affected region. Two types of spots can be noticed in the DR image. Spots of cotton wool (Figure 10 (a)) have yellowish-white, slightly raised lessons, and their fuzzy borders provide the illusion of clouds on the retina. The hard exudates (Figure 10 (b)) are tiny deposits that are white or pale yellow in color, with distinct edges. The resultant position of the identified test box is displayed via the bounding box regression. The impacted sections are used as a positive example for DR sign localization, whereas other portions and the background are used as a negative example. The overlapping region is labeled as background or harmful using the threshold value Intersection over Union (IoU), set to 0.3. If the IoU value is less than 0.3, the region is considered negative.

Similarly, regions with an IoU value greater than 0.7 are deemed to be the DR lesions. For the DR lesions localization, the Faster RCNN approach were used. As shown in Figure 10, the localization result of the Faster RCNN was applied to the dataset.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. EVALUATION CRITERION

In this exploration, DR detection performance was assessed by six metrics including accuracy (ACC), sensitivity (SE), specificity (SP), Precision (PR), mean average precision (mAP) and intersection over union (IoU), defined by Equation 12 to Equation 17. Here, TP stands for true positive, FP for false positive, TN for true negative, and FN for false negative.

$$Accuracy\,(ACC) = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Specificity\,(SP) = \frac{TN}{TN + FP} \tag{13}$$

$$Sensitivity\,(SE) = \frac{TP}{TP + FN} \tag{14}$$

$$Precision\,(PR) = \frac{TP}{TP + FP} \tag{15}$$

$$mAP = \sum_{i=1}^{T} AP(t_i)/T \tag{16}$$

$$IoU = \frac{Area\ of\ overlapp}{Area\ of\ union} \tag{17}$$

B. EXPERIMENTAL SETTINGS

This study used 21,171 retina FIs for training, testing, and validation without data augmentation to confirm the suggested system established for automated DR identification. To address the issue of data disequilibrium, the data distribution was laminated in this study. All data sets were separated into 0.8 and 0.2 portions for training and testing, respectively. The data distribution for developing the DR detection framework is presented in Table 2.

A PC with 64-bit Windows, 8 GB RAM, and an Intel Core i7 CPU with a processing speed of 3.00 GHz was used for all computations tasks. All relevant experiments were carried out using MATLAB 2019b. The experimenting times were calculated using a GEFORCE RTX 2070 super GPU configuration.

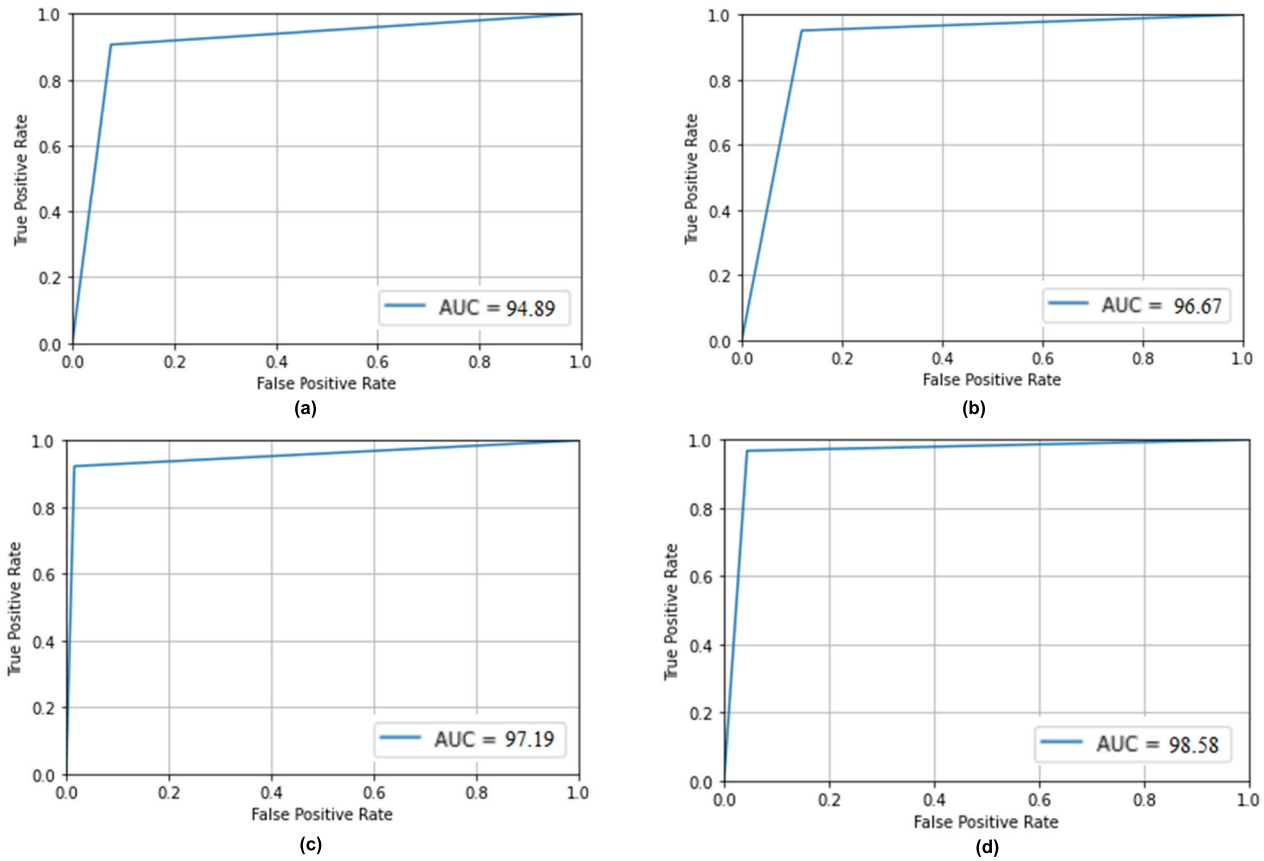


FIGURE 13. ROC curves of the overall framework with fused feature for (a) VGG19 (b) RCNN (b) fast RCNN (b) faster RCNN models.

TABLE 8. Performance of 5 sub fold cross-validation for test and training data.

Fold performance	Training accuracy	Validation accuracy	Training loss	Validation loss
Sub Fold-1	0.9667	0.9236	0.0522	0.129
Sub Fold-2	0.9589	0.9289	0.0675	0.152
Sub Fold-3	0.9597	0.9185	0.0785	0.189
Sub Fold-4	0.9712	0.9279	0.0678	0.275
Sub Fold-5	0.9745	0.9374	0.0734	0.179
Average	0.9662	0.9273	0.0678	0.1848

C. FILTERING METHOD PERFORMANCE

The research applied different filtering methods and the filtering performance was evaluated based on the criteria such as PSNR (Peak Signal to Noise Ratio), MSE (Mean Square Error) and Normalized Cross Correlation (NK) [27], [28]. Equation (18) to Equation (20) provide the MSE, PSNR and NK formulas. All of the filters are applied to the Gaussian-noise-affected retinal picture. The suggested filtering technique, Adaptive Median Filter (AMF), denoised the image better than the mean filter, average filter, and median filter as shown in Table 3. AMF produced the lowest mean square error, the highest peak signal to noise ratio, and

$$f_{HOG1\times n} = \{HOG_{1\times1}, HOG_{1\times2}, HOG_{1\times3}, \ldots\ldots\ldots\ldots .HOG_{1\times n} \quad (8)$$

$$f_{ST1\times m} = \{ST_{1\times1}, ST_{1\times2}, ST_{1\times3}, \ldots\ldots\ldots\ldots .ST_{1\times m}\} \quad (9)$$

$$f_{RCNN1\times p} = \{RCNN_{1\times1}, RCNN_{1\times2}, RCNN_{1\times3}, \ldots\ldots\ldots\ldots .RCNN_{1\times p}\} \quad (10)$$

$$Fused\ (featuresvector)^{cat}_{1\times q} = \{f_{HOG1\times n}, f_{ST1\times m}, f_{RCNN1\times p}\} \quad (11)$$

a content of 1.000 in Normalized Cross Correlation. This clearly illustrated that in the recovered image from the noisy image, the AMF produced the best quality.

$$MSE = \frac{1}{MN}\sum_{j=1}^{M}\sum_{k=1}^{N}(x_{j,k} - x_{j,k})^2 \tag{18}$$

$$PSNR = 10log\frac{(255) \wedge 2}{MSE} \tag{19}$$

$$NK = \sum_{j,k}^{M}\sum_{j,k}^{N} x_{j,k} - x'_{j,k'}\sum_{j,k}^{M}\sum_{j,k}^{N} x_{j,k}^2 \tag{20}$$

D. INDIVIDUAL AND FUSED FEATURE PERFORMANCES

In this section, features extracted by different methods were combined and faster RCNN classification technique was used to assess which feature fusion method produces the best performance. The neural features of the Faster RCNN model are additionally merged with the extracted features of HOG and Shearlet transform using the gradient descriptor. While concatenating with Faster RCNN, the fusion of HOG and Shear let transform exhibits superior classification results than the interpolation-oriented descriptor [40]. Table 4 shows the comparative analysis of the results obtained using individual feature extractors and different fused methods. The feature vector obtained by the fusion of three methods performed better than the other double fused methods.

Multiple CNN based feature extractors were merged with HOG and shearlet transform extractors and the fused feature performance was also investigated individually. Merging HOG and Shear let Transform features with the faster RCNN provided the best performance compared to the that meged with others CNN feature extraction methods such as LSTM, VGG16, ReNet50, VGG16, RCNN, Fast RCNN (Table 5).

E. CLASSIFICATION PERFORMANCES

After merging the three feature vector data into one vector, classification was performed using different RCNN classifier models along with VGG19 as shown in Table 6. The faster RCNN model again provided the best performance than the other RCNN or VGG19 with an accuracy of 98.58% and a specificity of 98.57%. Figure 11 shows different performance accuracy and loss curves for the classification. The fast RCNN validation accuracy curve (Figure 11(c)) significantly drops because some validation features might have deviated more than the training features for the outliners in the validation datasets.

Figure 12 shows the confusion matrix created from the entire system's performance measurement during the classification step. The suggested approach needed marked test data to evaluate its intended outcome during the evaluation. According to the proposed faster RCNN confusion matrix, out of 2,156 images, the model accurately detected 2,120 DR images and could not detect 36 DR images. Out of 2,077 healthy images, the system failed to detect 24 normal image and correctly detected 2,053 normal images.

Three two-stage detector models used were Fast-RCNN, Faster-RCNN, and RCNN. The main difference between the two models is that two-stage detectors work by first using different region proposal strategies to locate the important item in an image, which is then narrowed down before doing the final classification task. As mAP metric is very popular and considered a standard metric for solving subject identification problems, this was implemented to conduct the performance study of all object-detection models. Table 7 compares the findings, showing that our proposed method obtains 96% mAP, greater than the other methods. Furthermore, the test times of all of the models were measured to assess their computational complexity. The Faster-RCNN framework achieved the highest mAP value with the shortest test time, as shown by the reported results.

The Receiver Operating Characteristics (ROC) curve for the suggested approach is shown in Figure 13. Compared to the other methods, the Faster-RCNN approach produced a nearly 1.00 area under the ROC curve (AUC). This indicated that the developed technology was quite effective in detecting the DR from the retina FIs.

F. VALIDATION EVALUATION

Generalization validation techniques were used to validate the system's performance further. The term ''generalization'' refers to a model's ability to respond to new inputs. This technique can digest new data and produce accurate predictions after being taught. This study used the IDRID retina image dataset for generalization purposes, which was not the same as the training set [39] containing 516 images with two classes of DR (310 images) and normal images (206 images). From the confusion matrix results for the generalization outcomes, only 15 out of 310 DR images are incorrectly detected. The DR retina images had the lowest false detection rate, with a generalization accuracy of 95.34%. Even with a completely fresh set of data, the system's image classification reliability was demonstrated.

K-fold cross-validation is generally used in applied machine learning to evaluate a model's performance to generate predictions on data that was not utilized during the model's training [41]. To increase the confidence in the experimental outcomes, a five-fold cross-validation procedure was applied with the feature vector being randomly divided into 5 sub folds. Four sub folds were chosen for the training dataset, but for the testing dataset, just one sub fold was chosen. Table 8 shows the accuracy and loss values for different folds during training and validation. Figure 14 shows the performance of 5 fold cross validation for test and training data.

G. COMPARATIVE ANALYSIS

The mean performance evaluation metrics were obtained after carrying out 5 runs with randomly choosing the data 80:20 ratio for training and testing. Compared to other strategies reported in the literature, the proposed approach

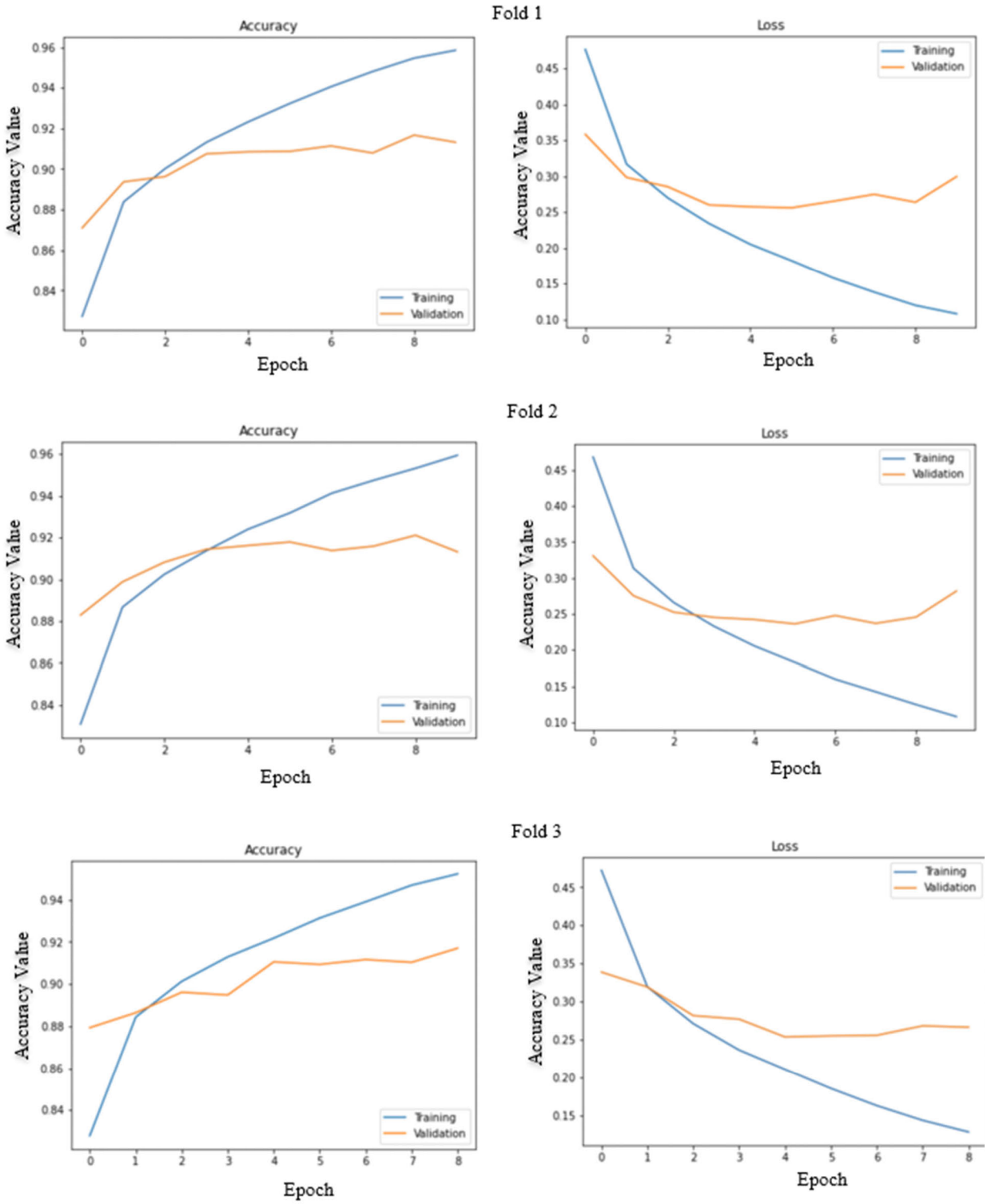


FIGURE 14. Accuracy and loss curves for 5-fold cross-validation study.

produced a higher classification accuracy score of 98.58% in detecting DR images (Table 9) possibly due to the utilization of fused features from three different extraction techniques.

Wan et al. [33], used large number of data (35,120 images) but overall performance was not as good (accuracy 95.68%). This could be due to the fact that the authors used a CNN based model to classify and extract features. In this model, a balanced dataset was employed to classify the DR images by using the fused features. Oliveira et al [35] found improved accuracy of 96.94% by combining multiscale

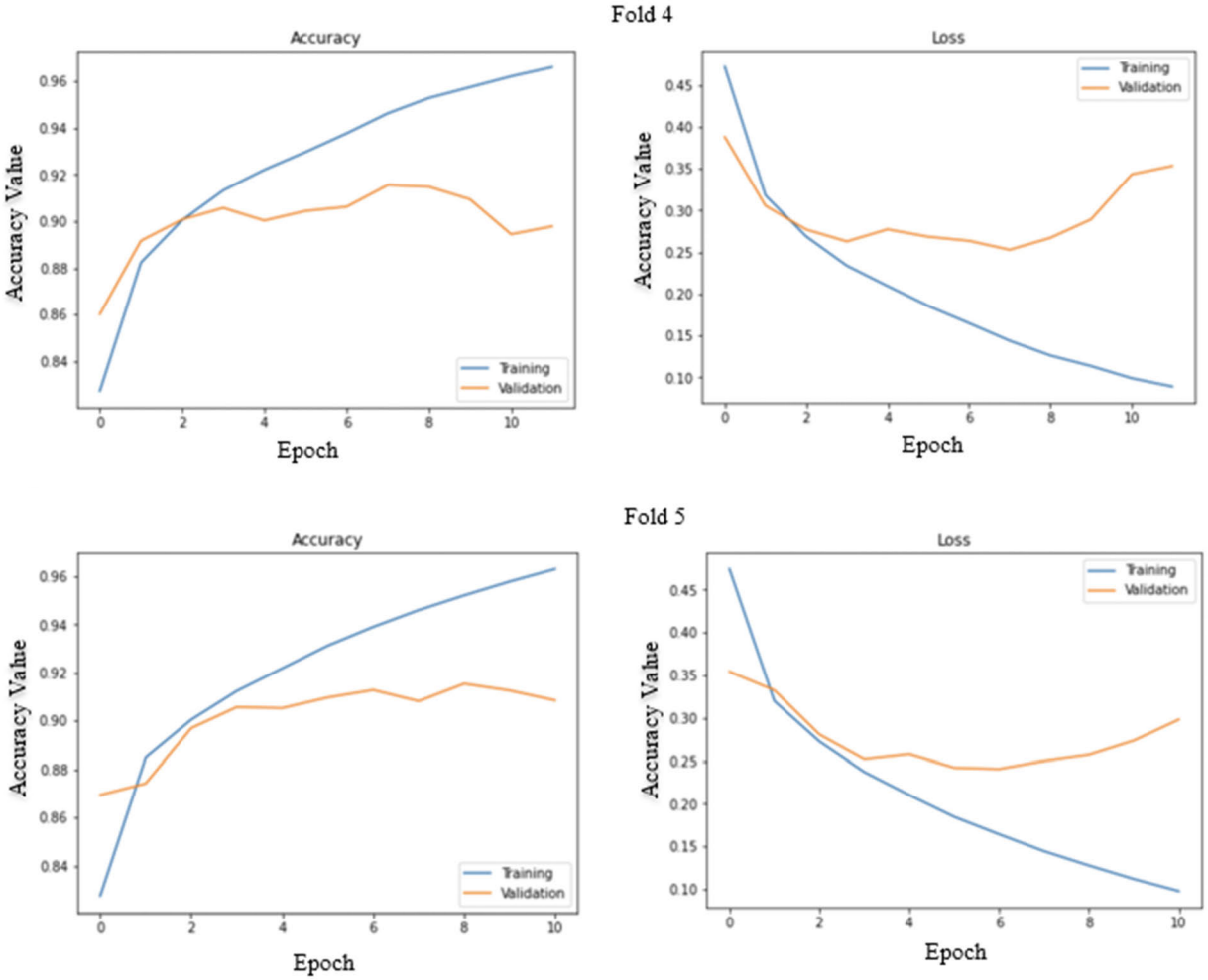


FIGURE 14. *(Continued.)* **Accuracy and loss curves for 5-fold cross-validation study.**

analysis provided by the Stationary Wavelet Transform with a multiscale Fully Convolutional Neural Network. This could be owing to the fact that they employed a significantly smaller dataset for testing the system's performance (total 88 images) [35]. In [36], the researcher only used the CNN model for feature extraction and classification and produced a competitive performance (97.89%) with less than half of the images in this dataset. Some studies [29], [30], [31], [32], [38] used small number of data ranging between 3,000 to 5,000 to achieve reasonable accuracy between 79 % to 95% using

CNN models. However, these models would take more time to run. Furthermore, it requires extensive training and has decreased accuracy when used to train imbalance samples. Islam et al. achieved an accuracy of 98.36% by using supervised contrastive learning (SCL) with supervised contrastive loss function [46]. The main drawbacks in that research identified are unbalanced datasets, grading errors, and resource limitations. According to the official APTOS 2019 documentation, the Messidor-2 dataset grading annotation contains flaws and was provided by a third party. The batch size was restricted to 8 due to the computational resource limitations. However, in the proposed model, application of balanced dataset could eliminate some of the limitations.

H. FALSE PREDICTION AND ANALYSIS

The system was tested with many poorly rated images to explore if it could properly detect them. The system did not provide the expected results for some unknown images collected from Kaggle [47] as shown in Figure 15 and also show why the images could not be detected by our system. These images were not used for training purposes.

I. LIMITATION AND FUTURE WORK

Despite the excellent performance achieved by the proposed method there is still room for improvement. For example, the DIARETDB1 dataset was imbalanced containing 84 DR images and 5 normal retinal images. This issue could be overcome by introducing generative adversarial network technique on both majority and minority classes. However, this did not make the overall dataset size highly imbalanced with over 21,000 images. Another limitation was that

TABLE 9. Comparative analysis of relevant research.

Ref.	DL method	Dataset	Number of Testing Image	Accuracy	Sensitivity	Specificity
[30]	CNN	Kaggle (5,000)	989	79.00%	90.00%	95.00%
[29]	CNN	Kaggle (1,000)	200	94.50%	92.30%	93.80%
[38]	CNN-ResNet34	Kaggle (3,000)	900	85.00%	86.00%	87.00%
[31]	DNN, CNN (VGGnet architecture), BNN	Kaggle (2,000)	350	78.39%	72.34%	73.78%
[32]	CNN (IntceptionNet v3)	Kaggle (166)	45	63.23%	62.67%	92.43%
[33]	CNN (AlexNet, VGG19, Googlenet and ResNet)	Kaggle (35,120)	3000	95.68%	90.78%	97.12%
[34]	Fully CNN	STARE (20), HRF (45), Drive (40)	21	96.28%	80.90%	97.70%
[35]	Fully CNN	STARE (20), HRF (45), Drive (40), CHASE_DB1 (28)	35	98.05%	97.67%	96.56%
[37]	CNN (ResNet50)	Messidor (1,200) and IDRID (516)	323	96.3%	92.6%	92%
[36]	CNN	Messidor 2 dataset (9,194)	2700	97.89%	91.83%	93.78%
[46]	Supervised contrastive learning	APTOS 2019 and Messidor-2 dataset	700	98.36%	98.37%	98.50%
-	**Proposed method (Faster RCNN+ Feature fusion)**	**DIARETDB1 (89), Kaggle (20,702), DDR (42)**	**4233**	**98.58%**	**98.57%**	**98.61%**

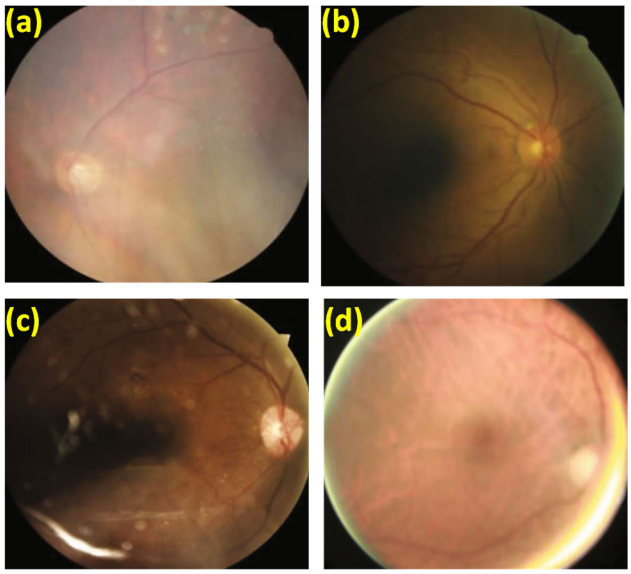


FIGURE 15. Images in the Kaggle dataset with wrong predictions: (a) False-negative in an image with foggy view and retinal laser scar, (b) False negative in an image with poor illumination, (c) False-negative in an image with reflective spots and shadows and (d) False-positive in an image with overexposure and halo.

the proposed feature fusion method misclassified 127 DR images. However, this could be solved by either training the model more or employing a more robust regularization strategy. Further fine tuning of the model is required to shorten the long training time resulted from the feature fusion operation. In the near future, it has been planned to collect diabetic macular edema images to enhance the model performance on a dataset containing diverse images. Furthermore, the effectiveness of the model needs to be tested in clinical set-up with even using mobile app.

IV. CONCLUSION

In this research, systematic attempts were undertaken to build an automated system that can forecast DR by analysing retina FIs from three different publicly available datasets and using fused features and Faster-RCNN classification. Findings suggests that the adaptive median filtering approaches provided better noise eliminations and obtain improved accuracy and precision from other filtering methods. The proposed model produced an accuracy, specificity, sensitivity and precision scores of 0.9858, 0.9757, 0.9861, 0.9859 during the binary classification of a modified dataset of over 21,000 retina images. According to the experimental results, all RCNN classification models achieved better results with Faster-RCNN classifier delivering the best performance with mAP, IoU, required time values of 0.965, 0.975, 0.20 s. It was also demonstrated that the proposed framework can precisely locate the DR affected region in the retina images. Furthermore, the model also generated state-of-the-art performance in comparison with the reported approaches in the literature.

REFERENCES

[1] *Global Action Plan on Physical Activity 2018–2030: More Active People for a Healthier World*, World Health Org., Geneva, Switzerland, 2019.

[2] R. Williams, S. Karuranga, B. Malanda, P. Saeedi, A. Basit, S. Besançon, C. Bommer, A. Esteghamati, K. Ogurtsova, P. Zhang, and S. Colagiuri, "Global and regional estimates and projections of diabetes-related health expenditure: Results from the international diabetes federation diabetes atlas, 9th edition," *Diabetes Res. Clin. Pract.*, vol. 162, Apr. 2020, Art. no. 108072.

[3] American Diabetes Association, "Diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 37, no. 1, pp. S81–S90, 2014.

[4] R. R. Bourne, G. A. Stevens, R. A. White, J. L. Smith, S. R. Flaxman, H. Price, J. B. Jonas, J. Keeffe, J. Leasher, K. Naidoo, K. Pesudovs, S. Resnikoff, and H. R. Taylor, "Causes of vision loss worldwide, 1990–2010: A systematic analysis," *Lancet Global Health*, vol. 1, no. 6, pp. 339–349, 2013.

[5] M. Ragab, A. S. A.-M. Al-Ghamdi, B. Fakieh, H. Choudhry, R. F. Mansour, and D. Koundal, "Prediction of diabetes through retinal images using deep neural network," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–6, Jun. 2022, doi: 10.1155/2022/7887908.

[6] J.-H. Wu, T. Y. A. Liu, W.-T. Hsu, J. H.-C. Ho, and C.-C. Lee, "Performance and limitation of machine learning algorithms for diabetic retinopathy screening: Meta-analysis," *J. Med. Internet Res.*, vol. 23, no. 7, Jul. 2021, Art. no. e23863, doi: 10.2196/23863.

[7] T. Nazir, A. Irtaza, A. Javed, H. Malik, D. Hussain, and R. A. Naqvi, "Retinal image analysis for diabetes-based eye disease detection using deep learning," *Appl. Sci.*, vol. 10, no. 18, p. 6185, Sep. 2020, doi: 10.3390/app10186185.

[8] S. S. Rahim, V. Palade, I. Almakky, and A. Holzinger, "Detection of diabetic retinopathy and maculopathy in eye fundus images using deep learning and image augmentation," in *Machine Learning and Knowledge Extraction* (Lecture Notes in Computer Science), vol. 11713, A. Holzinger, P. Kieseberg, A. Tjoa, and E. Weippl, Eds. Cham, Switzerland: Springer, 2019, doi: 10.1007/978-3-030-29726-8_8.

[9] M. Maniruzzaman, M. J. Rahman, B. Ahammed, and M. M. Abedin, "Classification and prediction of diabetes disease using machine learning paradigm," *Health Inf. Sci. Syst.*, vol. 8, no. 1, Dec. 2020, Art. no. 7, doi: 10.1007/s13755-019-0095-z.

[10] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, Dec. 2019, Art. no. 211, doi: 10.1186/s12911-019-0918-5.

[11] C. C. Kwan and A. A. Fawzi, "Imaging and biomarkers in diabetic macular edema and diabetic retinopathy," *Current Diabetes Rep.*, vol. 19, no. 10, pp. 1–10, Oct. 2019.

[12] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "DREAM: Diabetic retinopathy analysis using machine learning," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 5, pp. 1717–1728, Sep. 2014.

[13] L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet, and J. M. P. Langlois, "Red lesion detection using dynamic shape features for diabetic retinopathy screening," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1116–1126, Apr. 2016.

[14] D. Marin, M. E. Gegundez-Arias, B. Ponte, F. Alvarez, J. Garrido, C. Ortega, M. J. Vasallo, and J. M. Bravo, "An exudate detection method for diagnosis risk of diabetic macular edema in retinal images using feature-based and supervised classification," *Med. Biol. Eng. Comput.*, vol. 56, no. 8, pp. 1379–1390, Aug. 2018.

[15] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Ophthalmol. Vis. Sci.*, vol. 57, no. 13, pp. 5200–5206, 2016.

[16] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, Jul. 2017.

[17] D. Zhang, W. Bu, and X. Wu, "Diabetic retinopathy classification using deeply supervised ResNet," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Aug. 2017, pp. 1–6.

[18] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informat. Med. Unlocked*, vol. 20, Jan. 2020, Art. no. 100377, doi: 10.1016/j.imu.2020.100377.

[19] S. Albahli, T. Nazir, A. Irtaza, and A. Javed, "Recognition and detection of diabetic retinopathy using DenseNet-65 based faster-RCNN," *Comput., Mater. Continua*, vol. 67, no. 2, pp. 1333–1351, 2021.

[20] P. Samant and R. Agarwal, "Machine learning techniques for medical diagnosis of diabetes using iris images," *Comput. Methods Programs Biomed.*, vol. 157, pp. 121–128, Apr. 2018.

[21] G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. Teatini Salles, "Diabetic retinopathy detection using red lesion localization and convolutional neural networks," *Comput. Biol. Med.*, vol. 116, Jan. 2020, Art. no. 103537.

[22] J. Wu, Q. Zhang, M. Liu, Z. Xiao, F. Zhang, L. Geng, Y. Liu, and W. Wang, "Diabetic macular edema grading based on improved faster R-CNN and MD-ResNet," *Signal, Image Video Process.*, vol. 15, no. 4, pp. 743–751, Jun. 2021.

[23] X. Guo, X. Lu, Q. Liu, and X. Che, "EMFN: Enhanced multi-feature fusion network for hard exudate detection in fundus images," *IEEE Access*, vol. 7, pp. 176912–176920, 2019.

[24] X. Pan, K. Jin, J. Cao, Z. Liu, J. Wu, K. You, Y. Lu, Y. Xu, Z. Su, J. Jiang, K. Yao, and J. Ye, "Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 258, pp. 779–785, Jan. 2020.

[25] H. Riaz, J. Park, H. Choi, H. Kim, and J. Kim, "Deep and densely connected networks for classification of diabetic retinopathy," *Diagnostics*, vol. 10, no. 1, p. 24, Jan. 2020.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[27] D. Guo, Y. Wu, S. S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

[28] I. Aizenberg, C. Butakoff, and D. Paliy, "Impulsive noise removal using threshold Boolean filtering based on the impulse detecting functions," *IEEE Signal Process. Lett.*, vol. 12, no. 1, pp. 63–66, Jan. 2005.

[29] 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. 118164–118173, 2020.

[30] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, I.-H. Ra, and M. Alazab, "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, p. 274, Feb. 2020.

[31] S. Dutta, B. C. Manideep, S. M. Basha, R. D. Caytiles, and N. C. S. N. Iyengar, "Classification of diabetic retinopathy images by using deep learning models," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 1, pp. 99–106, Jan. 2018.

[32] X. Wang, Y. Lu, Y. Wang, and W.-B. Chen, "Diabetic retinopathy stage classification using convolutional neural networks," in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, Jul. 2018, pp. 465–471.

[33] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," *Comput. Electr. Eng.*, vol. 72, pp. 274–282, Nov. 2018.

[34] J. Lu, Y. Xu, M. Chen, and Y. Luo, "A coarse-to-fine fully convolutional neural network for fundus vessel segmentation," *Symmetry*, vol. 10, no. 11, p. 607, Nov. 2018.

[35] A. Oliveira, S. Pereira, and C. A. Silva, "Retinal vessel segmentation based on fully convolutional neural networks," *Expert Syst. Appl.*, vol. 112, pp. 229–242, Dec. 2018.

[36] J. Wang, J. Luo, B. Liu, R. Feng, L. Lu, and H. Zou, "Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm," *IET Comput. Vis.*, vol. 14, no. 1, pp. 1–8, Feb. 2020.

[37] X. Li, X. Hu, L. Yu, L. Zhu, C.-W. Fu, and P.-A. Heng, "CANet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1483–1493, May 2020.

[38] M. T. Esfahani, M. Ghaderi, and R. Kafiyeh, "Classification of diabetic and normal fundus images using new deep learning method," *Leonardo Electron. J. Pract. Technol.*, vol. 17, no. 32, pp. 233–248, 2018.

[39] *Academic Torrents*. Accessed: Aug. 10, 2021. [Online]. Available: https://academictorrents.com/

[40] N. Sasikala, V. Swathipriya, M. Ashwini, V. Preethi, A. Pranavi, and M. Ranjith, "Feature extraction of real-time image using SIFT algorithm," *Eur. J. Electr. Eng. Comput. Sci.*, vol. 4, no. 3, pp. 1–7, May 2020.

[41] M. M. Hasan, N. Islam, and M. M. Rahman, "Gastrointestinal polyp detection through a fusion of contourlet transform and neural features," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 34, no. 3, pp. 526–533, Mar. 2022.

[42] R. Mostafiz, M. S. Uddin, N.-A. Alam, M. M. Hasan, and M. M. Rahman, "MRI-based brain tumor detection using the fusion of histogram oriented gradients and neural features," *Evol. Intell.*, vol. 14, no. 2, pp. 1075–1087, Jun. 2021, doi: 10.1007/s12065-020-00550-1.

[43] T. Nazir, A. Irtaza, J. Rashid, M. Nawaz, and T. Mehmood, "Diabetic retinopathy lesions detection using faster-RCNN from retinal images," in *Proc. 1st Int. Conf. Smart Syst. Emerg. Technol. (SMARTTECH)*, Nov. 2020, pp. 38–42, doi: 10.1109/smart-tech49988.2020.00025.

[44] J. Du, B. Zou, C. Chen, Z. Xu, and Q. Liu, "Automatic microaneurysm detection in fundus image based on local cross-section transformation and multi-feature fusion," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105687.

[45] J. D. Bodapati, V. Naralasetti, S. N. Shareef, S. Hakak, M. Bilal, P. K. R. Maddikunta, and O. Jo, "Blended multi-modal deep ConvNet features for diabetic retinopathy severity prediction," *Electronics*, vol. 9, no. 6, p. 914, May 2020.

[46] 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," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105602, doi: 10.1016/j.compbiomed.2022.105602.

[47] Kaggle. (2015). *Diabetic Retinopathy Detection*. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection

MD. NUR-A-ALAM received the B.Sc. degree in engineering from the Department of Computing Science and Engineering, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh, in 2019, where he is currently pursuing the M.Sc. (Research) degree in engineering. He is currently an Academic Researcher and a Faculty Member with the Department of Computer Science and Engineering, Dhaka International University. He has published four research papers in different international journals and conferences. His research interests include image processing, machine learning, data mining, wireless communication, and deep learning. He is a member of the Bangladesh Computer Society.

MD. MOSTOFA KAMAL NASIR received the B.Sc. and M.Sc. degrees in computer science and engineering from Jahangirnagar University, Bangladesh, and the Ph.D. degree in mobile adhoc technology from the University of Malaya, Kuala Lumpur, Malaysia, in 2016. He is currently a Professor of computer science and engineering with the Mawlana Bhashani Science and Technology University, Tangail, Bangladesh. His current research interests include VANET, the IoT, SDN, and WSN.

MOMINUL AHSAN received the B.Sc. degree from the Department of Computer Science and Engineering, State University of Bangladesh, Dhaka, Bangladesh, the M.Eng. (Research) degree from the Faculty of Engineering and Computing, Dublin City University, Dublin, Ireland, and the Ph.D. degree from the School of Computing and Mathematical Sciences, University of Greenwich, London, U.K. He was a Postdoctoral Researcher with the Department of Engineering, Manchester Met University, before joining the University of York. He is currently an Associate Lecturer with the Department of Computer Science, University of York, U.K. He has published and presented more than 50 technical papers in international journals, conferences, and books. His research interests include prognostics, data analytics, artificial intelligence, machine learning, reliability, power electronics, wireless communication, and wearable technology. He is a member of the Institution of Engineering and Technology (MIET), U.K.; a fellow of the Higher Education Academy (FHEA), U.K.; and an Associate Member of the Bangladesh Computer Society. He was a recipient of the M.Eng. Stipend from Dublin City University, the Ph.D. VC Scholarship from the University of Greenwich, and the Excellent Poster Award from the IEEE International Spring Seminar on Electronic Technology, Sofia, Bulgaria, in 2017.

MD. ABDUL BASED (Member, IEEE) was born in Nilphamari, Bangladesh, in 1979. He received the B.S. degree in computer science and engineering from Dhaka International University, Dhaka, Bangladesh, in 2000, the first M.S. degree in computer science from North South University, in 2005, and the second M.S. degree in information and communication systems security (ICSS) from the Royal Institute of Technology (KTH), in 2008. From 2008 to 2014, he was a Research Fellow with the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, on the security aspects of electronic voting. Since 2014, he has been an Associate Professor with the Department of Electrical, Electronics and Telecommunication Engineering (EETE), Dhaka International University, where he has been the Head, since September. He was also the Head of the Department of Computer Science and Engineering, Dhaka International University. Besides teaching and research, he was an IT and MIS Specialist in a Government Project at the Ministry of Planning, Dhaka. He is an Expert in Curriculum Design and Development and Outcome-Based Education. His research interests include information security, electronic voting, and cryptographic protocols. He is a Life Time Member as well as an Elected Councilor of the Bangladesh Computer Society (BCS), Bangladesh, (2018–2020).

JULFIKAR HAIDER is currently a Senior Lecturer in mechanical engineering with the Department of Engineering, Manchester Metropolitan University, U.K. He has received funding from Innovate U.K. to conduct eight knowledge transfer partnership projects with the industry worth more than a million pounds. He has published and presented more than 150 technical papers in international journals, conferences, and books. His main research interests include materials processing, thin film coating, composite materials, finite element analysis, and artificial intelligence; and it is in these areas that he is renowned for the work both nationally and internationally justified by the research outputs. He has been supervising research associates and Ph.D. students in the above areas. He is acting as an Executive Editor for an international journal *Advances in Materials and Processing Technologies* (Taylor and Francis).

SIVAPRAKASAM PALANI received the B.E. degree in industrial engineering, the M.E. degree in computer-integrated manufacturing, and the Ph.D. degree in mechanical engineering from the College of Engineering, Guindy (CEG), Anna University, Chennai, India, in 2005, 2008, and 2015, respectively. He is currently an Associate Professor with Addis Ababa Science and Technology University, Ethiopia. He has 15 years of teaching and research experience. He has published 85 technical papers in international journals/conferences and book chapters. His research interests include micro/nano-manufacturing, machining, composite materials, surface engineering, and optimization techniques. He has been supervising master's and Ph.D. students in the above areas.