IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 26, 2021, accepted January 28, 2021, date of publication February 1, 2021, date of current version February 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056186

Diabetic Retinopathy Detection and Classification Using Mixed Models for a Disease Grading Database

ANAS BILAL, (Member, IEEE), GUANGMIN SUN, YU LI, SARAH MAZHAR, AND ABDUL QADIR KHAN

Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Corresponding author: Guangmin Sun (gmsun@bjut.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 11527801 and Grant 41706201.

ABSTRACT Diabetic retinopathy (DR) is a primary cause of blindness in which damage occurs to the retina due to an accretion of sugar levels in the blood. Therefore, prior detection, classification, and diagnosis of DR can prevent vision loss in diabetic patients. We proposed a novel and hybrid approach for prior DR detection and classification. We combined distinctive models to make the DR detection process robust or less error-prone while determining the classification based on the majority voting method. The proposed work follows preprocessing feature extraction and classification steps. The preprocessing step enhances abnormality presence as well as segmentation; the extraction step acquires merely relevant features; and the classification step uses classifiers such as support vector machine (SVM), K-nearest neighbor (KNN), and binary trees (BT). To accomplish this work, multiple severities of disease grading databases were used and achieved an accuracy of 98.06%, sensitivity of 83.67%, and 100% specificity.

INDEX TERMS Binary trees, diabetic retinopathy, feature extraction, KNN, majority voting system, multiclass classification, segmentation, SVM.

I. INTRODUCTION

T he Diabetic retinopathy (DR) is considered one of the prime causes of blindness due to retinal blood vessel distortion and scarring. Therefore, the prior detection and diagnosis of DR biomarkers tend to prevent vision loss in the early stages of diabetes [1], [2]. To diagnose DR, ophthalmologists follow a conventional fundus image analysis method where the detection of microaneurysms (MAs) is being observed. MAs are the primary sign of DR. However, such a traditional screening method has severe difficulties; for example, it is laborious, has a high error-prone probability (due to manual screening) and is not a cost-effective solution. In addition, misclassifications may occur when some elements in the fundus images are marked as MAs, while other faults occur when clustering MAs with different shapes on the same class [1]. Computer-aided automated diagnosis (CAD) methods can detect DR in a short time and offer a high precision rate. Therefore, conventional methods can be easily replaced where visual evaluation and observation are manually required [3]. DR has been classified into two levels: proliferative diabetic retinopathy (PDR) and nonproliferative diabetic retinopathy (NPDR), as depicted in Fig. 1. The PDR level (see Fig. 1F) indicates normal functionality of the retina and blood vessels and has a clear optic disk. The NPDR level is considered a primary stage of DR, and its anomalies are classified based on severity (mild, moderate, and severe) and are shown in Fig. 1 (C-E) where the small veins (blood vessels) in the retina start to leak blood or fluid (substances) into the retina; therefore, it causes the retina to store the exudates. Since the skin becomes moist and inflamed, disorders generally include hemorrhaging, exudate, and microaneurysms that can be detected at the NPDR level [4]. When vision is influenced, macular enema is also expected to collapse, which causes swelling or thickening in the retina. Indeed, it is a tiny area in the focal point of the retina that helps us to easily see subtle fine elements plainly [5]. However, several methods have been proposed and developed to classify early detection and severity of DR (mild, moderate, or severe). In the mild stage, each blood vessel has a small red dot with one MA at

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson.

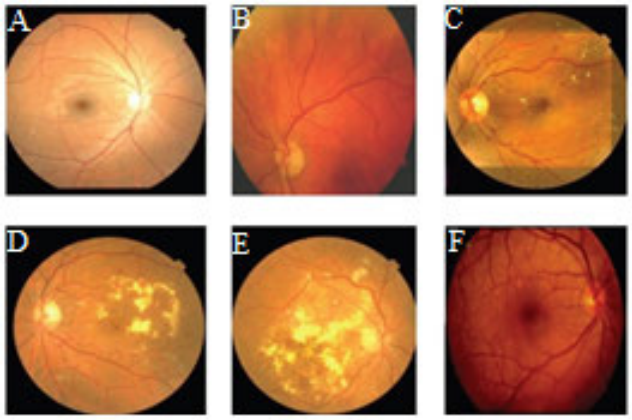


FIGURE 1. Fundus images of (A) normal retina, (B) early DR, (C) mild NPDR, (D) moderate NPDR, (E) severe NPDR, and (F) PDR [7].

the end of it, while the moderate stage occurs in the deeper layers of the retina and forms a hemorrhage; in the severe stage, the retina contains more than 20 hemorrhages [6].

The combination of image processing and data-mining techniques is commonly employed to examine fundus screening. In this process, image processing techniques are generally deployed to extract the retinal fundus image features. Data-mining techniques are applied to build a learning model for classification; this classifier can recognize the presence (or absence) of the disease in retinal fundus images [8]. The implemented CAD methods for DR early detection are familiar and useful and considered a major challenge since they are novel in the medical field. However, such methods deliver reliable and accurate disease diagnosis results [9]. Three machine-learning-based algorithms are initially deployed to classify the fundus images into three classes in the present work. Later, the classifiers' output is used as inputs to a voting method to obtain the final results based on each classifier's maximum number of votes.

This study's novelty combines distinctive models based on the majority voting method, which leads to less error proneness when deciding on classification. The combination of image enhancement study and feature extraction allows the detection process to perform a specific characterization of the DR.Similarly, combining a set of features from different regions and fusing them to construct a more extensive feature set increases the overall accuracy.

The rest of the article is organized as follows. A literature review and its related work are presented in section II. Section III presents a detailed description of the methodology. The results and discussions are given in Section IV, and the comments of the proposed work are presented in Section V.

II. RELATED WORK

In the context of hybrid methods, various authors have proposed unique hybrid methods for the DR detection process where disparate algorithms are used based on the dataset source, preprocessing, feature extraction, and algorithms for classification methods. [8], Usman and Khalid proposed a new technique for detecting microaneurysms in fundus images using genetic programming (GP), which is also called the intelligent feature set tuning (IFST) technique. They improved feature extraction and their numbers in the preclassification process by generating a mathematical expression for classifying MA images with the MESSIDOR and DIARETDB1 datasets. [9], Gayathri *et al.* proposed binary and multiclass DR classification where the Haralick and anisotropic dual-tree complex wavelet transform (ADTCWT) technique was used for feature extraction from fundus images. The authors chose the MESSIDOR, KAGGLE, and DIARETDB0 databases to obtain a reliable DR classification where classifiers such as the support vector machine (SVM), random forest, random tree, and J48 techniques were implemented. For classifying color retinal fundus images, Zeng *et al.* [10] proceeded with a deep learning algorithm-based CAD method to diagnose referable DR, while the Siamese-like architecture of the convolution neural network (CNN) model simultaneously accepted binocular fundus images to extract the characteristic features for Kaggle's (retina image) dataset and achieved a success rate of 95.1%.

A detailed review of DR detection was presented by Mateen *et al.* [11], where the retinal datasets, DR detection methods, and performance evaluation metrics were assessed. Moreover, the authors considered all available DR datasets and research development. Manjramkar [12], surveyed numerous available DR detection process techniques and studied the detection of exudates at an early stage to prevent vision loss. Therefore, he built a CAD model for DR early detection, which is useful for ophthalmologists. Qummar *et al.* [13], used Kaggle's dataset to train five deep CNN models, which improved the classification of robust features for DR severity. The training set was divided into tiny batches. In contrast, the CNN model was optimized to reduce empiric loss. GeethaRamani *et al.* [14], proposed a model in two classification levels for DR. On the first level, the best first tree set was performed, and the J48 graft tree set was used in the second using the clean data from the first classification level. They used the Drebechen dataset from the UCI machine learning repository (UCIMLR) and achieved an accuracy of 96.14%.

Ghazal *et al.* [15], devised a CNN-based CAD model of an optical coherence tomography (OCT) scanner for normal retinal appearance and detected NPDR. This model achieved an accuracy of 94%.A deep CNN system for semantic segmentation of fundus images was proposed by Qiao *et al.* [16] to evaluate the presence of a microaneurysm. The CNN model was considered a graphics processing unit (GPU)-accelerated core component that results in the detection and segmentation of medical images. The semantic segmentation algorithm has been implemented to split the image pixels based on their common semantics to identify the MA feature for classifying the fundus images as either normal or infected. Safi *et al.* [17], carried out a review to explain the studies on subclinical DR biomarker development and how they could be applied for

regular clinical use to identify potential perspectives for the detection of subclinical DR. Their review centered on a range of research methods and techniques for early DR detection. However, the authors confirmed that the current techniques would not deliver reliably predicted values.Olafsdottir *et al.* [18], perceived the differences in the severity of diabetes and diagnostic approaches by comparing screening and conventional methods. The authors assessed 257 diabetic patients' data, where 151 patients were detected through the screening process, and the rest were detected through the conventional process (clinical care). The fundus image study showed that only 22% of the patients were detected by screening (retinopathy), while 51% were clinically detected.

Soomro *et al.* [19]proposed a stride-CNN model for retinal vessel segmentation based on small vessels. The model is a fully convolutional one. The morphological mappings produce contrast images for training datasets and the Principal Component Analysis (PCA)-based pre-processing steps. We have shown that their model has a sensitivity of 0.9566, 0.954, 0.976, and 0.962.Soomro *et al.* [20]proposed a new segmentation method to address low sensitivity, using modules such as color-to-gray conversion, principal component analysis-based (PCA) scale normalization factors. For improved narrow vessel detection, anisotropic diffusion filtering with an adequate stopping rule and edge pixel-based hysteresis threshold were used. On publicly accessible databases such as DRIVE and STARE, the effect of these additional steps is measured. The sensitivity is increased from 73 to 75 percent for the DRIVE database, thus retaining 96.5 percent accuracy and providing evidence of enhanced sensitivity. An automated retinal vessel segmentation system is suggested for rapid and precise eye vessel segmentation [21]. Independent Component Analysis (ICA) is primarily used and consists of two architectures for noise reduction. On retinal color Fundus images, they validated both ICA architectures and select the one that provides enhanced contrast values. On publicly accessible databases, including DRIVE and STARE, the proposed segmentation model's effect was evaluated. The sensitivity increased 3 percent (from 72% to 75%) in the DRIVE database while retaining a segmentation precision of around 96%. Soomro *et al.* [22] propose a new algorithm for retinal vessel segmentation to resolve the low sensitivity problem. Along with pre-processing and post-processing, the proposed approach introduces a deep learning model. Preprocessing is used to address the issue of uneven illumination. They developed a full Convolutional Neural Network (CNN) and trained it to help observe fine vessels. Good segmented images, particularly for the detection of small vessels, are given by the proposed segmentation process. On the widely used publicly accessible databases: DRIVE and STARE databases, they have assessed their technique. With an accuracy of 94.7 percent, a higher sensitivity of 75% contributes to the proper detection of small vessels. The authors used a patch-based deep network approach to develop a lesion localization system. They aimed to reduce the model's sophistication while enhancing its efficiency [23]. For retinal image recognition, Harangi *et al.* [24] proposed an ensemble that organizes a convolutional neural network (CNN) and typical handcrafted characteristics into a single architecture. This method enables the CNN joint preparation and fine-tuning of handcrafted features' weights to have a final forecast. Their experimental findings focused on IDRiD, and classification accuracy was 90.1%. By investigating the internal interaction between diseases with only image-level tracking, the authors propose a novel cross-disease attention network (CANet) to distinguish DR and DME jointly [25].

III. METHODOLOGY

This section describes the proposed methodology for the detection and classification of DR. First, we describe the general design of the proposed method, which summarizes the complete process's training and testing phases. The descriptions of the different modules are then presented in the following sub-sections. The proposed method's algorithm is preprocessing, followed by extracting features and DR classification, as illustrated in Fig. 2.

A. PREPROCESSING

This step is targeted at enhancing the given fundus image where it performs adaptive histogram equalization and contrast stretching and is followed by a median filtering procedure. The preprocessing step is applied in the luminance, chroma blue C_B, and chroma red C_R color space of the fundus image. It, therefore, combines two of the chromatic color components of the RGB to the intensity (Y), which results in the YC_bC_r image (transformation of RGB image to YC_bC_r image) [26]. The Y component is separated from the YC_bC_r image to apply the median filtering process. After that, the contrast stretching and intensity normalization process can be applied. Later, the image is retrieved to RGB (as an inverse transform of YC_bC_r, as explained in the algorithm subsection.

B. FEATURE EXTRACTION

To accomplish the feature extraction step, the preprocessed image is now fed to the bright and red lesion detection algorithms, which perform the feature extraction process on the detected regions.

C. CLASSIFICATION

After extracting the features, the classification process is initiated using three classifiers, namely, SVM, KNN, and BT. Each method uses an adaptive threshold algorithm to enhance images with a sensitivity level of 0.85. Subsequently, the morphological operation is applied to filter out the noise from the output of the image.

D. THE ALGORITHM OF THE PROPOSED METHOD

The proposed work algorithm is followed by Fig.2 segmentation steps as follows (#).

Preprocessing step: This module defines the necessary preprocessing for the better detection of the MAs and to

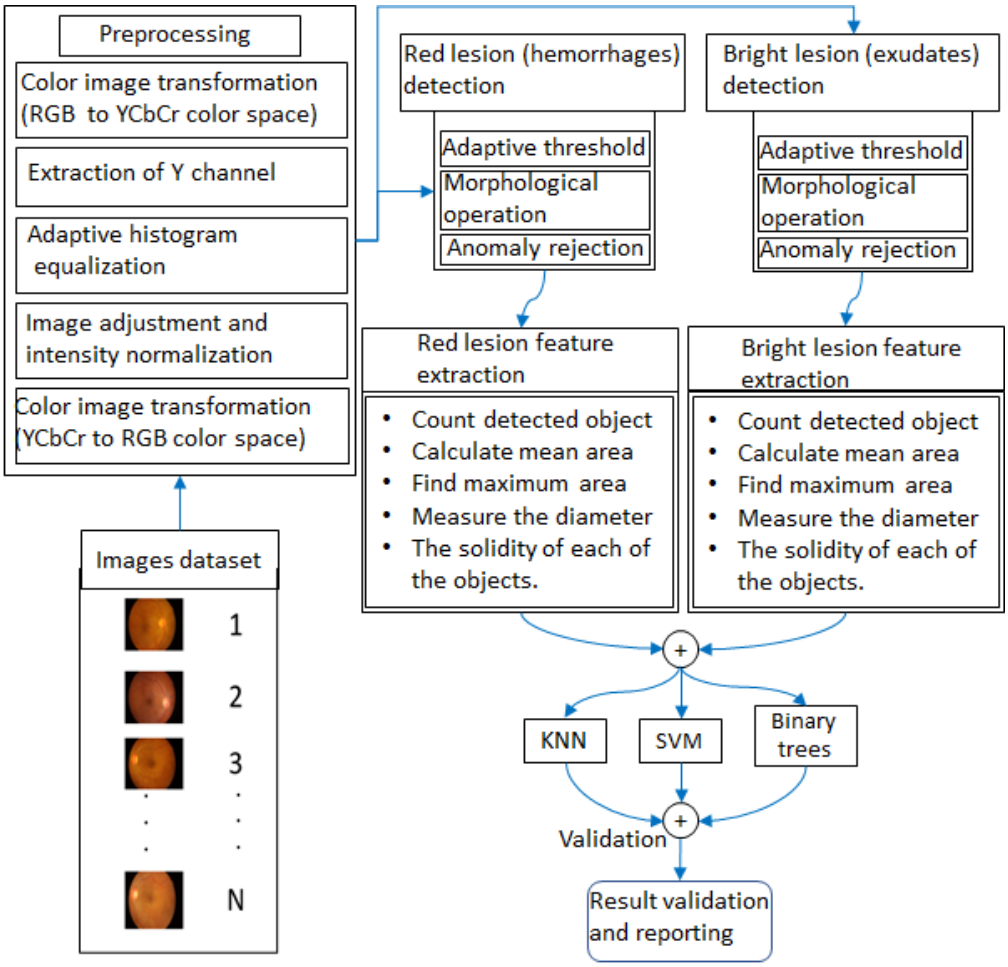


FIGURE 2. Block diagram of the proposed method.

remove the inherent and external noise induced in the fundus images during the creation and transmission process. The following is the sequence of steps we do follow:

- Convert the image from RGB to YC_bC_r and only take the Y component
- Apply a median filter
- Apply contrast stretching and intensity normalization
- Recover the YC_bC_r
- Convert the YC_bC_r back to the RGB.

Detection of the red lesion: This segment describes all the necessary processing steps for the detection of the hemorrhages and the following sequence of steps, we follow:

- Apply adaptive segmentation with a sensitivity value of 0.15 and dark foreground polarity
- Obtain the segmented properties, extends, and aspect ratio
- Filter out the segmented image as per extent and aspect ratio

Red lesion feature extraction: This segment describes all necessary processing steps for feature extraction of the red lesion and the following sequence of steps, we follow:

- Obtain all-region properties
- Obtain the number of regions # 1st feature
- Obtain the mean area of all regions # 2nd feature
- Obtain the mean perimeter of all regions # 3rd feature
- Obtain mean solidity of all regions # 4th feature
- Stack all the features together.

Detection of bright lesion: This section describes all the necessary processing steps for the detection of the exudates in the fundus image and the following sequence of steps, we follow:

- Apply adaptive segmentation with a sensitivity value of 0.85 and bright foreground polarity.
- Apply area filtering and morphological closing.

Bright lesions features: This segment is quite similar to the red lesion feature extraction process and following sequence of steps, we follow:

- Obtain all-region properties,
- Obtain the number of regions # 1st feature,
- Obtain the mean area of all regions # 2nd feature,
- Obtain the mean perimeter of all regions # 3rd feature,
- Obtain mean solidity of all regions #4th feature.

Fusion: This segmentation step is employed after the red and the white lesion feature extraction process is completed.

- The red lesion features and bright lesion features are appended lexicographically.

Training algorithm: This segment describes the essential steps of the image construction (training) and set of targets and the following sequence of steps, we follow:

- For i = 1, to the number of training images,
- Obtain the ith image from the database,
- Apply to preprocess and extract the features,
- Stack the feature results to the training image array,
- Assign a target class based on the severity of the dataset.

Train classifiers: This sub-segment defines the required steps to train the selected classifier and the following classifiers are chosen to perform the operations,

- Train the SVM classifier,
- Train the KNN classifier, and
- Train the BT classifier.

Testing algorithm: This segment defines the overall processing steps to predict the results through the feature extraction process by the given classifier and the following sequence of steps, we follow:

- Obtain the desired test image,
- Apply image preprocessing and extract the features,
- Predict the SVM classifier features,
- Predict the BT classifier features,
- Predict the KNN classifier features,
- Obtain the model of three prediction results.

The bright lesion detection algorithm is presented in Fig. 2, where the morphological operation is accomplished through the adaptive threshold and filtering process. The other region that the algorithm attempts to detect is the red lesions, which contain low-intensity values (relatively) after the preprocessing step. Nevertheless, the red lesions and bright algorithm are similar apart from sensitivity.

1) An adaptive threshold algorithm
2) Morphological operations
3) Anomaly rejection algorithm.

The red lesion sensitivity adaptive threshold is less than 0.15, whereas the morphological operations (bright regions) are approximately the same. The anomaly rejection algorithm can filter out the extracted vein (presented in the red regions) with an aspect ratio threshold and extend the threshold to measure the tested object to be filled on its corresponding bounding box. Finally, the following features are calibrated to make the classification:

1) Count the detected objects.
2) The mean area, maximum area, diameter, and solidity of each detected object were calculated.

It is worth mentioning that the algorithm outputs have three sets of features, namely, red, bright, and fused features of red and bright regions ordered lexicographically in the vertical direction. However, the fused features are the features selected to perform the rest of the classification. To accomplish the classification task, three classifiers are trained: SVM, KNN, and BT. A voting method is followed to obtain the final output (maximum number of votes for each classifier output). The classification accounted for the following 5 severities:

TABLE 1. KNN Parameters

Fitness Function	$\sqrt{(\sum (p_i - q_i)^2)}$
Distance metric	City block distance
K (number of neighbors)	1
Distance weight	Equal
Data normalization	$y = (x - \mu)/\sigma$ Where, μ and σ are the mean and standard deviation of each feature, respectively.
Multi-class coding style	one vs. all

1) No DR
2) Mild DR
3) Moderate DR
4) Severe DR
5) Proliferative DR.

Each model's accuracy is reported as the number of precisely classified labels over the total number of images per class.

E. TRAINING AND TESTING

The training and testing are described in Fig.2. The training process is initiated by preprocessing the whole training, and then the features are extracted from the exudates and hemorrhage regions. All the training features obtained from all the images are fed into the three classifiers individually by the target classes. Last, the classifiers are saved for testing.Likewise, the testing process is also initiated by the preprocessing step. In addition, each classifier's prediction values are considered a vote, and the mode of votes is calculated. Later, classification can be decided by the higher vote.

The parameters settings and the fitness functions for the utilized classifiers are discussed below:

1) K-NEAREST NEIGHBORS

The most affecting factors that affect tree performance is the number of neighbors and the distance metric. For the distance metric, the city block distance was used, and the number of neighbors was "1," which takes first the nearest neighbor as the output class. As presented in Table 1, the key parameter settings are held the same.

2) SUPPORT VECTOR MACHINE

The support vector is applied with a radial basis function Gaussian kernel with data normalization before feeding it to the SVM. Finally, the regularization parameter was set to "1," referring to no regularization during training. The key parameter settings in table 2 are preserved in the same way.

3) DECISION TREES

The main algorithm of the binary is aimed to produce maximally balanced trees. The binary trees split all the nodes in the current layer to satisfy the maximum number of splits and then count the number of branch nodes. A layer is a set of nodes equidistant from the root node. If the number of branch

TABLE 2. Support Vector Machine Parameters

Fitness Function	$\arg min(\frac{1}{n} * \sum_{i=1}^{N} \max(0, 1 - y_i(x_i.w - b)) + \|\lambda w^2\|) regularization\ parameter$
Learnable model parameters	Typically initialized [-1,1] w,b
Regularization parameter lemda	1
Kernel function	Radial basis function(RBF)
Data normalization	$y = (x - \mu)/\sigma$ Where, μ and σ are the mean and standard deviation of each feature, respectively.
Multi-class coding style	one vs. all

TABLE 3. Decision Tree Parameters

Maximum number of splits	1
Minimum parent size	10
Minimum leaf size	1
Data normalization	$y = (x - \mu)/\sigma$ Where, μ and σ are the mean and standard deviation of each feature, respectively.
Multi-class coding style	one vs. all

nodes reaches the maximum number of splits, it is followed by the algorithm. The configurations of the key parameters, as seen in Table 3 , are held the same.

1) Count how many branch nodes in the current layer need to be merged to be at most the maximum number of splits branch nodes.
2) Sort the branch nodes by their impurity gains.
3) Merge the desired number of least successful branches.
4) Output the decision tree grown so far.
   - The algorithm continues to split branch nodes layer by layer until at least one of this stopping criterion happens
   - The maximum number of splits is reached.
   - One of the splits at a specific iteration causes the number of observations in a particular branch less than the minimum parent size.
   - One of the splits in a particular iteration causes several observations in an individual leaf less than the minimum leaf size.

The decision trees, as mentioned it has the following hyper-parameters as mentioned in the methodology section above: The maximum number of splits value is set to 1.Minimum parent size is which controls the minimum number of branches in a particular node observation value is 10. Minimum leaf size, which is the minimum number of leaf node observations value is set to 1.

F. PERFORMANCE METRICS

The accuracy, specificity, sensitivity, and F1-score metrics are considered to assess the study's overall performance [27], [28]. Each model's accuracy is obtained as the number of exactly classified labels over the total number of images per class and is derived from Eq. (1) to Eq. (6).

$$Sensitivity(SEN) = \frac{TP}{TP + FN} \tag{1}$$

$$Specificity(SPE) = \frac{TN}{TN + FP} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

where

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

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

where T_P is the number of true cases, T_N is the number of true negative cases, F_P is the number of false-positive cases and F_N is the number of false-negative cases.

IV. RESULTS AND DISCUSSION

All of the performed experiments were done in a Matlab programming environment. We used an Intel Core i7 7th generation processor with 1TB SSD memory and 16 GB of RAM. The key findings of the image preprocessing, time complexity, and classifier results are emphasized in this section. Additionally, a comparison of the proposed work against conventional methods is presented.

A. INDIAN DIABETIC RETINOPATHY IMAGE DATASET

We used the Indian diabetic retinopathy image dataset (IDRiD) to assess this proposed work, which is officially the first dataset for India. The IDRiD dataset was built from actual clinical examinations at the Nanded (M.S.), eye clinic in India, and publicly available. These images were captured using Kowa VX-10α fundus camera with an emphasis on the macula. Pupils in both subjects were dilated by reducing tropicamide at a concentration of 0.5% before image acquisition. The final dataset is the collection of 516 images, which have 50° fields of view, a resolution of 4288 × 2848 pixels, and are stored in jpg format. It is the only dataset reflecting the common lesions of diabetic retinopathy and normal retinal structures marked up to the pixel level. These data provide an advantage for creating and evaluating DR image analysis algorithms for prior detection [29]. The dataset consists of five DR (0-4) and three DME (0-3) groups with well-defined characteristics in compliance with international clinical relevance standards. A grading score varies from 0 to 4, which represents the seriousness of the disease. There is no DR with a grade 0 patient. Grade 1, grade 2, and grade 3 patients have mild, moderate and severe non-proliferative DR levels. On the other side, grade 4 describes the most severe condition of DR. In the integrated dataset, 168 images (32.6%) are grade 0, 25 images (4.8%) are grade 1, 168 (32.6%) are grade 2,93 images (18%) are grade 3, and 62 (12%) are grade 4. In this study, we use 80% images for training purpose and

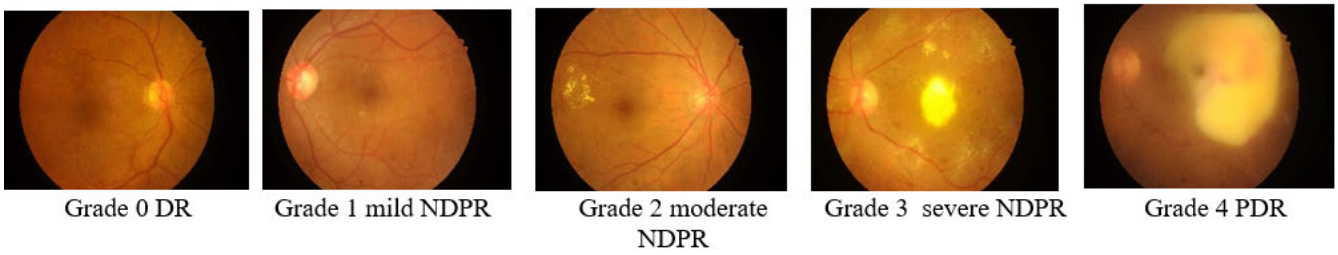


FIGURE 3. Sample images from the IDRiD dataset corresponding to disease severities.

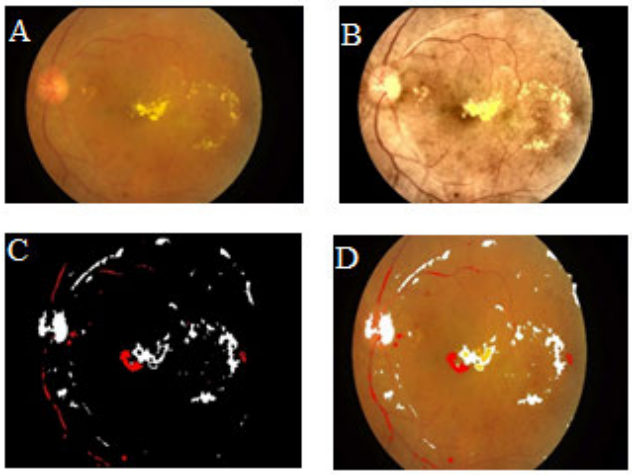


FIGURE 4. Image processing steps: (A) Original DR image, (B) enhanced image, (C) detected region (bright and red), and (D) segmentation results.

20% for testing with a ratio of 8:2, while a set of samples are presented in Fig. 3.

B. IMAGE PROCESSING RESULTS

A comparison of the preprocessing step image results and the classification final output results are presented in this section. A case study for disease grade 4 is examined, and the results are successfully obtained in the detection regions, as shown in Fig. 4. Since the disease grading database here does not provide any ground truth for the segmentation, no metrics are quantified for the segmentation result.

A comparison of different classifier results is presented in Table 4. Notably, the mixed model performs better than individual models. Among all classifiers, SVM yields better results than the BT model, whereas the KNN classifier yields the least accuracy. The results for higher disease severity generally represent better performance. The clearer disease appearance at high severities makes the features appear more discriminant. Therefore, it increases the classification process performance. The classification results of different disease severities and two performance metrics, namely, the accuracy and F1-score of the employed classification models, are depicted in Fig. 5. The mixed models (dashed red line) have shown better performance than other (SVM, KNN, and BT) classifiers. A comparison of the Vs class is presented in Table 5 through a confusion matrix. However, each entity in this table is the percentile of the total number of samples in the test set. The overall prediction rate for the mixed model's classification is 92.3%.

TABLE 4. Various Classification Results Score in this Work

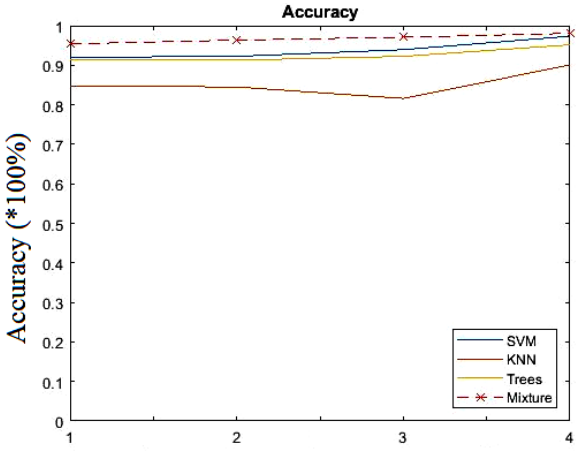
	Severity Threshold(≥)	Accuracy	Sensitivity	Specificity	F_1-Score
SVM	1	92.00	91.20	93.28	9392
	2	92.25	93.82	89.61	9382
	3	93.95	80.48	99.65	8878
	4	97.33	77.55	100.00	8735
KNN	1	84.75	85.30	83.58	8831
	2	84.50	86.10	81.81	8745
	3	81.59	57.72	91.72	6513
	4	90.07	42.85	96.42	5060
Binary Trees	1	91.28	91.39	91.04	9340
	2	91.28	92.66	88.96	9302
	3	92.25	82.92	96.20	8644
	4	95.15	71.42	98.35	7777
Mixed Models	**1**	**95.39**	**94.98**	**96.27**	**9563**
	2	**96.36**	**97.29**	**94.80**	**9710**
	3	**97.09**	**91.06**	**99.65**	**9491**
	4	**98.06**	**83.67**	**100.00**	**9111**

TABLE 5. The Confusion Matrix for Class Vs Class Comparison

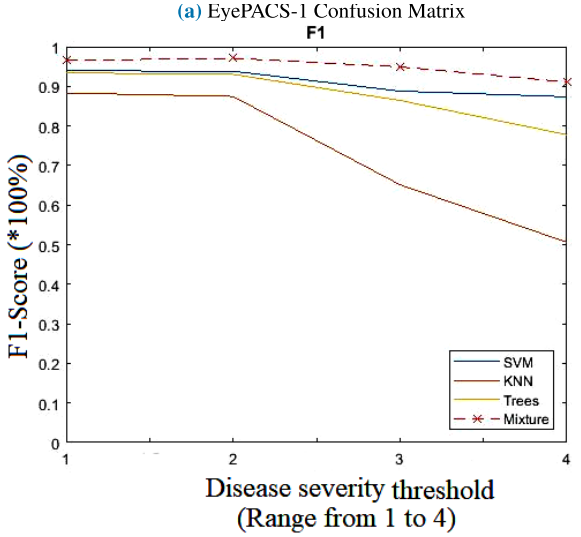
		Target Classes(%)					
Classifier output	Disease Grade	0	1	2	3	4	
	0	**32.1**	1.9	0.7	0.5	0.2	90.2
	1	0.0	**2.2**	0.2	0.0	0.0	90.0
	2	1.0	0.7	**32.0**	0.5	1.6	89.8
	3	0.2	0.0	0.0	**16.9**	0.2	97.2
	4	0.0	0.0	0.0	0.0	**9.0**	100
		96.3	45.0	97.1	94.6	83.7	**92.3**

C. TIME COMPLEXITY ANALYSIS

The processing time is a vital parameter in the image retrieval process, and the estimated (this work) processing time is presented in Table 6. The entire time procedure for each image is a combination of processing, training, and testing times. The processing time consists of preprocessing and feature extraction steps, where the process starts with reading the image and ends with feature extraction. Similarly, the time that it takes to train the entire dataset for each classifier is considered the training time. The testing time merely consists of the prediction and voting of each classifier. However, this study achieved an estimated processing time of approximately 10 seconds, which is relatively faster.



(a) EyePACS-1 Confusion Matrix



(b) Messidor-2 Confusion Matrix

FIGURE 5. The trends of (A) accuracy, (B) F1-score to the classifier types with different severities.

TABLE 6. The Entire Image Processing Time of the Study

Item	Time complexity(Sec)	Other info
Processing time(Preprocessing + Feature extraction)	9.563560	Per image
Training time for the three classifiers	0.437288	Per dataset
Testing time	0.037923	Per image

D. COMPARISON WITH THE STATE-OF-THE-ART STUDIES

Table 7 compares the proposed work vs conventional studies where the authors presented different databases to report their studies. Therefore, the parameters used in this comparison are a trade-off where few studies only consider the accuracy and sensitivity ([19]–[22], [30]–[32]) parameters, and some studies merely consider accuracy ([23]–[25]). The proposed work achieved better results in terms of accuracy and specificity. However, few studies have achieved a better sensitivity than the proposed work.

TABLE 7. A Comparison Between the Proposed Work and the State of the Art

Author	Accuracy (%)	Sensitivity (%)	Specificity (%)
Usman et al. [8]	97.00	98.00	96.00
Zeng et al. [10]	95.00	95.00	95.00
Qummar et al. [13]	80.80	86.70	63.80
Ramani et al. [14]	96.14	92.15	80.31
Ghazal et al. [15]	94.00	100	88.00
Soomro et al. [30]	95.90	80.20	-
Soomro et al. [19]	95.60	87.00	-
Soomro et al. [21]	96.00	75.00	-
Soomro et al. [20]	96.50	75.00	-
Soomro et al. [31]	94.70	75.00	-
Soomro et al. [32]	94.80	73.90	-
Soomro et al. [22]	94.70	75.00	-
Soomro et al. [33]	95.20	74.60	96.60
Soomro et al. [34]	94.40	-	-
Khan et al. [35]	95.01	73.66	96.89
Zago et l. [23]	89.10	-	-
Harangi et al. [24]	90.10	-	-
Li et al. [25]	92.60	-	-
Proposed work	**98.06**	**83.67**	**100**

E. CRITICAL ANALYSIS

Fig. 4 (A, B, C, and D) indicates that the preprocessing algorithm provides good performance in highlighting the regions containing the lesions. Therefore, two different algorithms (bright and red) are used to detect the (hemorrhages) lesions. Once the lesions are detected, the features are extracted independently and combined in a feature vector. The classification is performed depending on the increasing severity threshold by considering the group of metrics for each classifier, and the overall performance of the voting system is stored. It is noted that the voting system has the highest classification accuracy when compared to each classifier independently. In addition, a higher severity threshold is determined to increase all the classifiers' accuracies since lesion clarity makes them easy to classify. For the overall performance, the highest accuracy is achieved for mixed models (98.063%) at a disease severity of 4. In contrast, the highest F1-score and sensitivity were observed in the mixed models with a disease severity of 2. The time complexity performance of this algorithm is obtained to be relatively fast. The preprocessing procedure is observed as the most time-consuming process, with a duration of 9.5935 Sec, since all image processing steps are involved. However, the testing time to classify the image is relatively fast (0.0379 Sec).

V. CONCLUSION

This article presents the DR of the disease grading database. The methodology deployed in this work follows preprocessing feature extraction and classification steps and realizes three classifiers along with a combined voting method. In comparison, the proposed work achieved better results than the state of the art in terms of accuracy and specificity 98.06% and 100%, respectively. Nevertheless, the highest sensitivity (100%) is achieved in [15]. However, the achieved results are always a trade-off among the required parameters since the entire procedure is strictly dependent on the preprocessing

and feature extraction process. The three classifiers with the voting method strengthen the features that improve reliability.

Future studies may include implanting the algorithm with deep learning methodologies to compare the results with this current work.

REFERENCES

[1] H. Ahsan, "Diabetic retinopathy–biomolecules and multiple pathophysiology," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 9, no. 1, pp. 51–54, Jan. 2015.
[2] S. A. Schellini, G. M. D. Carvalho, F. S. Rendeiro, C. R. Padovani, and F. E. Hirai, "Prevalence of diabetes and diabetic retinopathy in a Brazilian population," *Ophthalmic Epidemiol.*, vol. 21, no. 1, pp. 33–38, Feb. 2014.
[3] P. G. Shekelle, S. C. Morton, and E. B. Keeler, "Costs and benefits of health information technology," *Evidence Rep. Technol. Assessment (Full Rep.)*, vol. 132, no. 4, pp. 1–71, 2006.
[4] C. P. Wilkinson, F. L. Ferris, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, and J. T. Verdaguer, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, Sep. 2003.
[5] J. D. M. Gass, A. Agarwal, A. M. Lavina, and K. A. Tawansy, "Focal inner retinal hemorrhages in patients with drusen: An early sign of occult choroidal neovascularization and chorioretinal anastomosis," *Retina*, vol. 23, no. 6, pp. 741–751, Dec. 2003.
[6] M. D. Abramoff, J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan, M. Niemeijer, and G. Quellec, "Automated early detection of diabetic retinopathy," *Ophthalmology*, vol. 117, no. 6, pp. 1147–1154, Jun. 2010.
[7] M. F. El-Bab, N. Shawki, A. AL-Sisi, and M. Akhtar, "Retinopathy and risk factors in diabetic patients from Al-Madinah Al-Munawarah in the kingdom of Saudi Arabia," *Clin. Ophthalmol.*, vol. 6, p. 269, Feb. 2012.
[8] I. Usman and K. A. Almejalli, "Intelligent automated detection of microaneurysms in fundus images using feature-set tuning," *IEEE Access*, vol. 8, pp. 65187–65196, 2020.
[9] S. Gayathri, A. K. Krishna, V. P. Gopi, and P. Palanisamy, "Automated binary and multiclass classification of diabetic retinopathy using Haralick and multiresolution features," *IEEE Access*, vol. 8, pp. 57497–57504, 2020.
[10] X. Zeng, H. Chen, Y. Luo, and W. Ye, "Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30744–30753, 2019.
[11] M. Mateen, J. Wen, M. Hassan, N. Nasrullah, S. Sun, and S. Hayat, "Automatic detection of diabetic retinopathy: A review on datasets, methods and evaluation metrics," *IEEE Access*, vol. 8, pp. 48784–48811, 2020.
[12] M. Manjramkar, "Survey of diabetic retinopathy screening methods," in *Proc. 2nd Int. Conf. Trends Electron. Informat. (ICOEI)*, May 2018, pp. 1–6.
[13] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. Ahmed Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019.
[14] R. G. Ramani, J. Shanthamalar J., and B. Lakshmi, "Automatic diabetic retinopathy detection through ensemble classification techniques automated diabetic retionapthy classification," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, Dec. 2017, pp. 1–4.
[15] M. Ghazal, S. S. Ali, A. H. Mahmoud, A. M. Shalaby, and A. El-Baz, "Accurate detection of non-proliferative diabetic retinopathy in optical coherence tomography images using convolutional neural networks," *IEEE Access*, vol. 8, pp. 34387–34397, 2020.
[16] L. Qiao, Y. Zhu, and H. Zhou, "Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms," *IEEE Access*, vol. 8, pp. 104292–104302, 2020.
[17] H. Safi, S. Safi, A. Hafezi-Moghadam, and H. Ahmadieh, "Early detection of diabetic retinopathy," *Surv. Ophthalmol.*, vol. 63, no. 5, pp. 601–608, 2018.
[18] E. Olafsdottir, D. K. G. Andersson, I. Dedorsson, K. Svärdsudd, S. P. O. Jansson, and E. Stefánsson, "Early detection of type 2 diabetes mellitus and screening for retinopathy are associated with reduced prevalence and severity of retinopathy," *Acta Ophthalmologica*, vol. 94, no. 3, pp. 232–239, May 2016.
[19] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, L. Zheng, and M. Paul, "Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation," *Expert Syst. Appl.*, vol. 134, pp. 36–52, Nov. 2019.
[20] T. A. Soomro, M. A. U. Khan, J. Gao, T. M. Khan, and M. Paul, "Contrast normalization steps for increased sensitivity of a retinal image segmentation method," *Signal, Image Video Process.*, vol. 11, no. 8, pp. 1509–1517, Nov. 2017.
[21] T. A. Soomro, T. M. Khan, M. A. U. Khan, J. Gao, M. Paul, and L. Zheng, "Impact of ICA-based image enhancement technique on retinal blood vessels segmentation," *IEEE Access*, vol. 6, pp. 3524–3538, 2018.
[22] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, M. A. U. Khan, M. Paul, and L. Zheng, "Boosting sensitivity of a retinal vessel segmentation algorithm with convolutional neural network," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2017, pp. 1–8.
[23] G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. T. Salles, "Diabetic retinopathy detection using red lesion localization and convolutional neural networks," *Comput. Biol. Med.*, vol. 116, Jan. 2020, Art. no. 103537.
[24] B. Harangi, J. Toth, A. Baran, and A. Hajdu, "Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2699–2702.
[25] 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.
[26] T. A. Soomro, J. Gao, T. Khan, A. F. M. Hani, M. A. U. Khan, and M. Paul, "Computerised approaches for the detection of diabetic retinopathy using retinal fundus images: A survey," *Pattern Anal. Appl.*, vol. 20, no. 4, pp. 927–961, Nov. 2017.
[27] F. E. Grubbs, "Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments," *Technometrics*, vol. 15, no. 1, pp. 53–66, Feb. 1973.
[28] M. Sharma, S. Sharma, and G. Singh, "Performance analysis of statistical and supervised learning techniques in stock data mining," *Data*, vol. 3, no. 4, p. 54, Nov. 2018.
[29] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, Jul. 2018.
[30] T. A. Soomro, A. J. Afifi, A. A. Shah, S. Soomro, G. A. Baloch, L. Zheng, M. Yin, and J. Gao, "Impact of image enhancement technique on CNN model for retinal blood vessels segmentation," *IEEE Access*, vol. 7, pp. 158183–158197, 2019.
[31] M. A. U. Khan, T. M. Khan, T. A. Soomro, N. Mir, and J. Gao, "Boosting sensitivity of a retinal vessel segmentation algorithm," *Pattern Anal. Appl.*, vol. 22, no. 2, pp. 583–599, May 2019.
[32] T. A. Soomro, A. J. Afifi, J. Gao, O. Hellwich, M. Paul, and L. Zheng, "Strided U-Net model: Retinal vessels segmentation using dice loss," in *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, Dec. 2018, pp. 1–8.
[33] T. A. Soomro, M. Paul, J. Gao, and L. Zheng, "Retinal blood vessel extraction method based on basic filtering schemes," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2017, pp. 4422–4426.
[34] T. A. Soomro, M. A. U. Khan, J. Gao, T. M. Khan, M. Paul, and N. Mir, "Automatic retinal vessel extraction algorithm," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2016, pp. 1–8.
[35] M. A. U. Khan, T. A. Soomro, T. M. Khan, D. G. Bailey, J. Gao, and N. Mir, "Automatic retinal vessel extraction algorithm based on contrast-sensitive schemes," in *Proc. Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, Nov. 2016, pp. 1–5.

ANAS BILAL (Member, IEEE) received the B.S. degree in telecommunication and networks from Iqra University, Pakistan, in 2013, and the M.S. degree in electrical and electronic systems from the University of Lahore, Pakistan, in 2016. He is currently pursuing the Ph.D. degree with the School of Information Technology, Beijing University of Technology. His research interests include neural networks, medical image analysis, machine learning, sentiment analysis, pattern recognition, and signal & image processing.

GUANGMIN SUN received the B.Sc. degree in electronic engineering from the Beijing Institute of Technology, China, in 1982, the M.S. degree in communication and information systems from the Nanjing University of Science and Technology, China, in 1991, and the Ph.D. degree in communication and information systems from Xidian University, China. He is currently a Professor with the Department of Electronic Engineering, Beijing University of Technology. His research interests include neural networks and applications, pattern recognition, bioinformatics, and signal and image processing.

YU LI received the B.S. degree in electronic and information engineering and the M.E. degree in information and communication engineering from Beihang University, Beijing, China, in 2009 and 2012, respectively, and the Ph.D. degree in the earth system and geo-information science (microwave remote sensing) from The Chinese University of Hong Kong, Hong Kong, China, in 2015. He is currently an Associate Professor of information and communication engineering with the Beijing University of Technology, Beijing. His research interests include image and signal processing, pattern recognition, and remote sensing applications.

SARAH MAZHAR received the M.S. degree in computer science from the National University of Modern Languages, Islamabad, Pakistan. She is currently pursuing the Ph.D. degree with the Beijing University of Technology, Beijing, China. She is also a Lecturer of computer science with the Faculty of Engineering and Computer Science, National University of Modern Languages. Her research interests include neural networks, image processing, remote sensing, sustainable systematic literature reviews, emotion-based requirement engineering, green computing, fog computing, and agile practices.

ABDUL QADIR KHAN received the B.Tech. degree (Hons.) in electronic technology from Preston University, Pakistan, in 2014, and the M.S.E.E.S. degree in electronics and electrical systems from the University of Lahore, Pakistan, in 2016. He is currently pursuing the Ph.D. degree with the School of Information, Beijing University of Technology. His research interests include antennas, cellular mobile communication, cellular network generation, networking, and machine learning.