Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 3, March 2020

A UNIQUE AUTOMATIC DIABETIC RETINOPATHY DETECTION SYSTEM USING MIXED MODELS THAT ACHIEVES 98% ACCURACY

[1]Manasa Thorrikon Da, [2]Sujatha Kaipa, [3]Naveen Kumar Polu, [4]Vadapalli Prasanna
[1, 2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Electronics and Communication Engineering, Rishi MS Institute of Engineering and Technlogy for Women, Kukatpally, Hyderabad.

ABSTRACT
Diabetes is the primary cause of the eye disorder known as diabetic retinopathy (DR). If the high glucose levels aren't under control, this condition gradually impairs the retina. Non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) are the two forms (PDR). In this study, we propose an accurate automated technique for classifying and identifying the various DR phases. The intended work follows the pre-processing, feature extraction, and feature classification methods. The pre-processing stage enhances the existence of anomalies in addition to segmentation. Only relevant characteristics are gathered throughout the extraction process. Support vector machines (SVM), K-nearest neighbours (KNN), and binary trees are used in the classification stage (BT). Additionally employed for classification is the random forest classifier, and the outcomes are contrasted with those from the other classifiers. To accomplish this, multiple severities of disease grading databases were used and achieved an accuracy of 98.06%, sensitivity of 83.67%, and 100% specificity.
Keywords*:* Support Vector Machine (SVM), K-nearest neighbour (KNN), Binary Tree (BT), Random Forest classifier (RFC), Database (Drive)

1. INTRODUCTION

Diabetic retinopathy is an eye disorder that mostly affects people with diabetes. This fatal ailment has affected over 290 million people worldwide, including 69.2 million individuals in India. The number of victims will increase dramatically in the ensuing years. The sickness is connected to the area of emphasis and, if left undiagnosed, might have detrimental effects on the patient. When the retina's blood vessels leak fluid into the retina, exudates occurs, which leads to the development of diabetic retinopathy? As a result, the diabetic patient can lose their vision. Micro aneurysms, exudates, and hemorrhages are what cause vision loss in persons with DR. As the retinal blood vessels expand, the vision loss increases.
The following are the photographs of healthy retina and diabetic retinopathy retina.

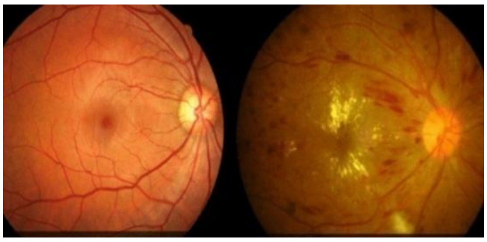
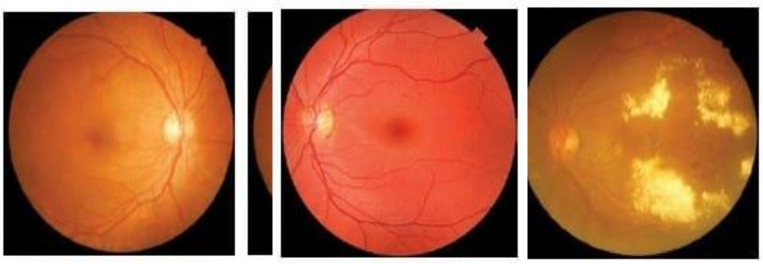


Fig: a. Healthy Retina   b. Diabetic Retinopathy Retina

There are two stages of DR :
Proliferative diabetic retinopathy (PDR) and non-proliferative diabetic retinopathy (NPDR)
The PDR level indicates normal functionality of the retina and blood vessels and features a clear optic disc. The NPDR level is taken into account as a primary stage of DR, and its anomalies are classified supported

severity (mild, moderate, and severe) where the little veins (blood vessels) within 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 hemorrhages, exudates, and micro aneurysms.

Stages of diabetic retinopathy:



a.) No DR b.) Mild DR c.) Moderate DR d.) Severe DR

A fast-approaching medical attention can be brought by early identification of the disease. Automatic screening of those images would help the doctors to simply detect the patient's condition in additional accurate way. Therefore, conventional methods are often easily replaced where visual valuation and observation are manually required. During this paper ,we emphasized on determination of retinal images using appropriate image processing techniques. With this we will easily classify normal and abnormal images of retina this may reduce the quantity of reviews for the doctors.

2. RELATED WORK

The International Clinical Diabetic Retinopathy (ICDR) scale is that the generally utilized clinical scales and comprises of 5-Point grade for Diabetic Retinopathy as an example typical, gentle, moderate, serious and proliferative. This categorization proposed by ICDR employed by many AI algorithms to understand the severity of Diabetic Retinopathy (DR).

Recent surveys are saying that just about 40 million of individuals in India are suffering from diabetes and thanks to this cause, most of the people are littered with DR. 90% of the DR patients are often saved by early diagnosis. The prevention and identification of DR may be done manually or automatically. Manual process could be a time taking process. During this process ophthalmologists must have an expertise within the work. In present scenario because of the advancement within the ophthalmology, AI is giving expertise to identifying and reducing the numerous diseases including Retinopathy caused by diabetes creator suggested the solution for reviewing DR and diabetic macular edema (DME) together.

The automated detection of microaneurysms and exudates was conducted on to 2 small image databases, where lesions were marked manually. The computer- assisted diagnostic system for the grading and detection of diabetic retinopathy and macular edema (ME) risks was improved, employing an oversized database including both pathological and normal images and differentiate manual gradings.

The proposed method failed to achieve a high accuracy, but it led to the reduction of the consumption time when using large amounts of knowledge. The presence of dark and bright lesions is an unsolved problem within the detection of blood vessels in retina images. within the detection of dark and bright lesions, the segmentation of blood vessels has a very important role within the processing step.

With the goal of prevailing the unevenly distributed noise, Lam B.S et al. designed the locally normalized concavity measure in keeping with modifications of spherical intensity. Also, for robust vessels segmentation, the perceptive of Weber’s law was accustomed map input images. The achieved results indicated a good performance of suggested methods on both normal and abnormal retina images.

3. METHODOLOGY:

The methodology follows a less complicated approach than the present one.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 3, March 2020

The detection of diabetic retinopathy is one among the various applications of the pattern recognition approaches. Pattern recognition may be considered as a classification task. Pattern recognition provides useful information supported observations. to style a pattern recognition system, we'd like to map from the measurement space into a meaningful space, where different points have different meanings. The fundamental components of a pattern recognition system are pre-processing, feature extraction and selection, classifier design and optimization.

Hence, our proposed diabetic retinopathy recognition approach consists of three main steps: pre-processing, feature extraction and have classification followed by identification of diabetic retinopathy. The diagram is as follows.

A detail information is discussed below:

PREPROCESSING

The pre-processing step aims at improving the given fundus image. It performs the subsequent methods like adaptive histogram equalization, contrast stretching followed by a median filtering procedure.

The primary step in pre-processing is colour image transformation. The given input retina image is in RGB format. The RGB colour space consists of 2^24 colour components during which it's substantially complicate to analyse all the colour components. So, we intended to Pre-processing step which is applied within the luminance, chrominance blue CB, and chrominance red CR colour space of the fundus image.

It combines two of the spectral colour components of the RGB to the intensity (Y), which ends up within the YCbCr image (transformation of RGB image to YCbCr image) . The Y component is extracted from the YCbCr image to use the median filtering process. After that, the contrast stretching and intensity normalization process are often applied.

Later, the image is retrieved to RGB (as an inverse transform of YCbCr , as explained within the algorithm subsection.

FEATURE EXTRACTION

Exudates, microaneurysms, haemorrhages will be extracted from retina image on the premise of color, intensity and texture. We try to extract these relevant and significant features from the fundus image to differentiate them. The feature extraction step is accomplished by feeding the pre-processed image to the intense and red lesion detection algorithms, which perform the feature extraction process on the detected regions. The feature extraction process is performed as said above.

FEATURE CLASSIFICATION:

The feature classification process is initiated using the classifiers like SVM, BT, KNN, RFC after extracting the features from detecting algorithms. the mixture of image processing and machine learning techniques is often employed to look at fundus screening.

During this process, image processing techniques are generally deployed to extract the retinal fundus image features. Machine learning techniques are applied to make a learning model for classification; this classifier can recognize the presence (or absence) of the disease in retinal fundus images. Three machine-learning-based algorithms are initially deployed to classify the fundus images into three classes within the present work.

Later, the classifiers' output is employed as inputs to a voting method to get the ultimate results supported each classifier's maximum number of votes.

THE ALGORITHM OF THE PROPOSED METHOD

The proposed work algorithm is as follows:

Pre-processing step:

This is the first step in the proposed work. In this step, it detects the MAs and it removes the inherent and external noise induced in the funds images during the creation and transmission process. The following is the sequences procedure of steps we do follow:

· Transform the image from RGB to YCbCr and only take the Y component
· Pertain median filter,
· Contrast stretching and intensity normalization
· Followedby Recovering the YCbCr
· Finally, reconstruct the YCbCr back to the RGB.

Discernment of the red lesion:

This segment describes all the necessary processing steps for the detection of the haemorrhages and the following sequence of steps, we follow:

· Pertain adaptive segmentation with a sensitivity value of0.15 and dark foreground polarity
· Acquire 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
· Acquire the number of regions of 1st feature, the mean area of all regions in 2nd feature, the mean perimeter of all regions in 3rd feature, the mean solidity of all regions in 4th feature
· And Finally, Stack all the features together.

Discernment 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:

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

Bright lesions feature:

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

· Acquire all-region properties,
· Acquire the number of regions in 1st feature, the mean area of all regions in 2nd feature, the mean perimeter of all regions in 3rd feature, and the mean solidity of all regions in 4th feature.

Fusion:

The last portion in the pre-processing step is Fusion. This segmentation step is employed after the red and the white lesion feature extraction process is completed.

Classifiers Description:

Support Vector Machine:

Support Vector Machine (SVM) is an algorithm that was developed for pattern classification but has recently been adapted for other uses, like finding regression and distribution estimation. It been utilized in many fields like bioinformatics, and is currently a really active research area in many universities and research institutes which include the National University of Singapore (NUS) and Massachusetts Institute of Technology (MIT). Although the SVM is applied to varied optimization problems like regression, the classic problem is that of knowledge classification. The essential idea is discussed. The info points are identified as being positive or negative, and therefore the problem is to search out a hyper-plane that separates the info points by a maximal margin.

Data Classification KNN Classifier:

In statistics, the k-nearest neighbor's algorithm (k-NN) could be a non-parametric supervised learning

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 3, March 2020

method first developed by Evelyn Fix and Joseph Hodges in 1951, and later expanded by Thomas Cover. It's used for classification and regression. In both cases, the input consists of the k closest training examples in a very data set. The output depends on whether k-NN is employed for classification or regression: • in k-NN classification, the output could be a class membership. So it is classed by a plurality vote of its neighbors, with the item being assigned to the category commonest among its k nearest neighbors (k may be a positive integer, typically small). If k = 1, then the item is solely assigned to the category of that single nearest neighbor. • In k-NN regression, the output is that the property value for the item. This value is that the average values of k nearest neighbors.

Binary Tree:

A binary tree may be a tree-type non-linear organization with a maximum of two children for every parent. Every node in a very binary tree encompasses a left and right reference together with the info element. The node at the highest of the hierarchy of a tree is termed the foundation node. The nodes that hold other sub-nodes are the parent nodes. The complete binary tree is additionally called a strict binary tree. The tree can only be considered because the full binary tree if each node must contain either 0 or 2 children. The total binary tree can even be defined because the trees within which each node must contain 2 children except the leaf nodes.

Random Forest Algorithm:

The Random Forest Algorithm consists of various decision trees, each with the identical nodes, but using different data that results in different leaves. It merges the selections of multiple decision trees so as to seek out a solution, which represents the common of these decision trees. The random forest algorithm could be a supervised learning model; it uses labeled data to "learn" the way to classify unlabelled data. this is often the alternative of the K-means Cluster algorithm, which we learned in an exceedingly past article was an unsupervised learning model. The Random Forest Algorithm is employed to resolve both regression and classification problems, making it a various model that's widely employed by engineers.

PERFORMANCE METRICS includes: accuracy, specificity, sensitivity, and F1-score metrics are considered to assess the study's overall performance. Each model's accuracy is obtained because the number of exactly classified labels over the overall number of images per class and springs from Eq. (1) to Eq. (6).

$$Sensitivity(SEN) = TP/TP + FN \qquad (1)$$

$$Specificity(SPE) = TN/TN + FP \qquad (2)$$

$$Accuracy = TP + TN/TP + TN + FP + FN \qquad (3)$$

$$F1 - Score = 2 * Precision * Recall\ Precision + Recall \qquad (4)$$

$$where\ Precision = TP/\ TP + FP \qquad (5)$$

$$Recall = TP/TP + FN \qquad (6)$$

Where

TP is that the number of true cases, TN is that the number of true negative cases, FP is that the number of false-positive cases and FN is that the number of false-negative cases.

TRAINING AND TESTING

The training process is initiated by pre-processing the entire training, and so the features are extracted from the exudates and hemorrhage regions. All the training features obtained from all the photographs are fed into the three classifiers individually by the target classes. Last, the classifiers are saved for testing.

Likewise, the testing process is additionally initiated by the pre-processing step. Additionally, each classifier's prediction values are considered a vote, and also the mode of votes is calculated. Later, classification is decided by the upper vote. The parameters settings and therefore the fitness functions for the utilized classifiers are discussed. The storing of feature vectors and class labels of the training samples are possible at the training stage. Frequently used distance metric for continuous variables is Euclidean distance.

For discrete variables, like for text classification, another metric are often used, like the overlap metric (or

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 3, March 2020

Hamming distance). Within the context of organic phenomenon microarray data, as an example, k-NN has been employed with correlation coefficients, like Pearson and Spearman, as a metric. A drawback of the essential "majority voting" classification occurs when the category distribution is skewed. That is, samples of a more frequent class tend to dominate the prediction of the new example, because they have a tendency to be common among the k nearest neighbors because of their sizable amount. a technique to beat this problem is to weight the classification, taking under consideration the space from the test point to every of its k nearest neighbors. The category (or value, in regression problems) of every of the k nearest points is multiplied by a weight proportional to the inverse of the gap from that time to the test point. in a different way to beat skew is by abstraction in data representation.

For instance, in a self-organizing map (SOM), each node may be a representative (a centre) of a cluster of comparable points, irrespective of their density within the original training data.

KNN can then be applied to the SOM.

Training algorithm

This segment describes the essential steps of the image construction (training) and set of targets and therefore the following sequence of steps, we follow: · For i D 1, to the quantity of coaching images, · Obtain the it image from the database, · Apply to pre-process and extract the features, · Stack the feature results to the training image array, · Assign a target class supported the severity of the dataset. Train classifiers: This sub-segment defines the specified steps to coach the chosen classifier and therefore 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 general processing steps to predict the results through the feature extraction process by the given classifier and therefore the following sequence of steps, we follow: · Obtain the specified test image, · Apply image pre-processing 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 where the morphological operation is accomplished through the adaptive threshold and filtering process. The opposite region that the algorithm attempts to detect is that the red lesions, which contain low-intensity values (relatively) after the pre-processing step. Nevertheless, the red lesions and bright algorithm are similar except for sensitivity. The anomaly rejection algorithm can separate out the extracted vein (presented within the red regions) with a side ratio threshold and extend the brink to live the tested object to be filled on its corresponding bounding box.

Finally, the subsequent features are calibrated to create the classification:

1) Count the detected objects.

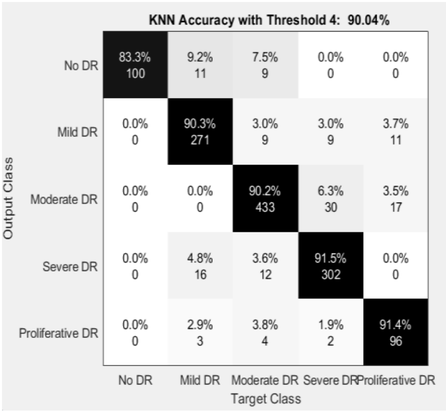
2) The mean area, maximum area, diameter, and solidity of every detected object were calculated.

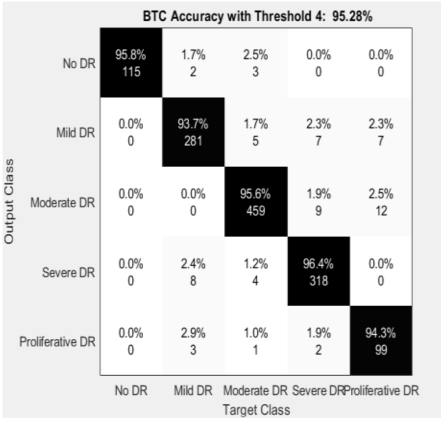
It is worth mentioning that the algorithm outputs have three sets of features, namely, red, bright, and fused features of red and bright regions ordered lexicographically within the vertical direction. However, the fused features are the features selected to perform the remainder of the classification. 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 every classifier output). The classification accounted for the subsequent 5 severities: 1) No DR 2) Mild DR 3) Moderate DR 4) SevereDR5) Proliferative DR. Each model's accuracy is reported because the number of precisely classified labels over the full number of images per class.

Experimental Analysis

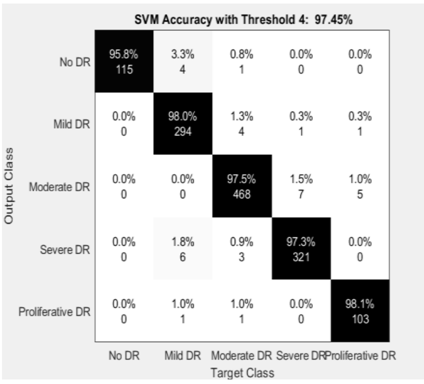
The following are the confusion matrices which show the experimental analysis of each of the classifiers.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 3, March 2020

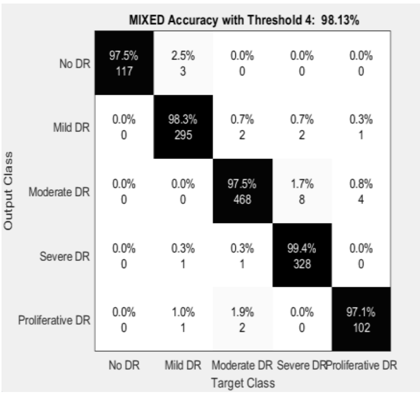


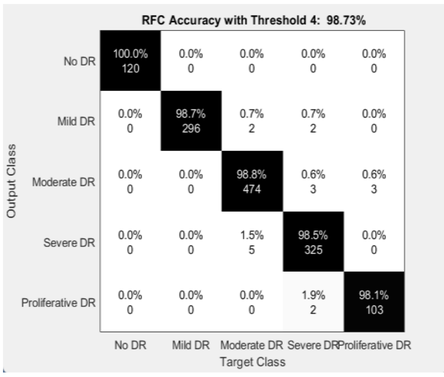


KNN MATRIX **BTC MATRIX**



SVM MATRIX



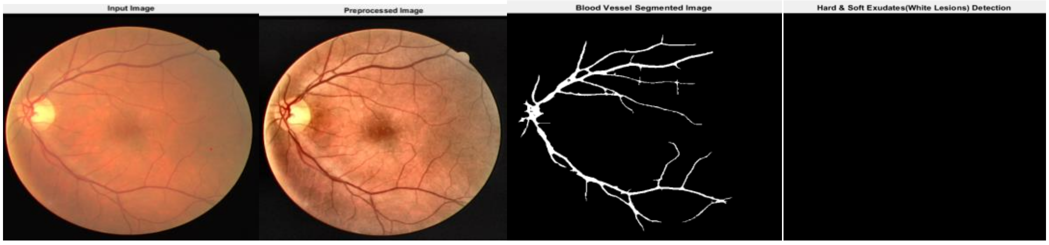


MIXED MODEL MATRIX **RFC MATRIX**

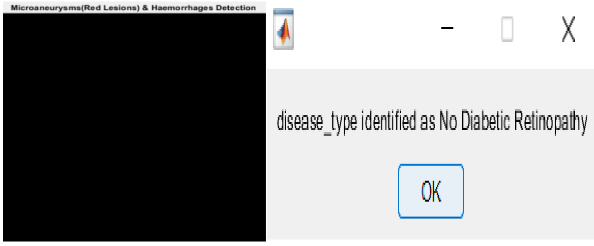
RESULTS AND DISCUSSION:

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

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 3, March 2020



Input image preprocessed image Blood vessel segment detection Hard and Soft exudates image detection



Micro aneurysms and Hemorrhages detection Output

CONCLUSION

The time it takes to make a diagnosis is significantly reduced by computerized systems, saving ophthalmologists' time and money while also facilitating quick patient treatment. For detecting DR at an early stage, automated DR detection systems are crucial. The DR stages are supported by the types of lesions that form on the retina. This article evaluates the most recent machine learning-based automated classification methods for diabetic retinopathy. Common funds and machine learning techniques are briefly discussed. Publicly available datasets like STARE and DRIVE have been described. This study has also discussed the useful techniques that will be used to categories and identify DR using ML algorithms. These machine learning methods are used for the prediction of Diabetic Retinopathy in patients. With the help of ML algorithms, knowledge is extracted for the prediction of DR. Four sorts of machine learning models were accustomed learn the results. This text will bring convenience together with low cost and high accuracy, as compared to traditional DR methods.

REFERENCES

[1] S. A. Schellini, G. M. D. Carvalho, F. S. Rendeiro, C. R. Padovani, andF. E. Hirai, ''Prevalence of diabetes and diabetic retinopathy in a Brazilianpopulation,'' Ophthalmic Epidemiol., vol. 21, no. 1, pp. 33–38, Feb. 2014.

[2] P. G. Shekelle, S. C. Morton, and E. B. Keeler, ''Costs and benefits of healthinformation technology,'' Evidence Rep. Technol. Assessment (Full Rep.),vol. 132, no. 4, pp. 1–71, 2006.

[3] 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, ''Proposedinternational clinical diabetic retinopathy and diabetic macular edemadisease severity scales,'' Ophthalmology, vol. 110, no. 9, pp. 1677–1682,Sep. 2003.

[4] M. D. Abramoff, J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan,M. Niemeijer, and G. Quellec, ''Automated early detection of diabeticretinopathy,'' Ophthalmology, vol. 117, no. 6, pp. 1147–1154, Jun. 2010.

[5] 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.

[6] S. Gayathri, A. K. Krishna, V. P. Gopi, and P. Palanisamy, ''Automatedbinary and multiclass classification of diabetic retinopathy using Haralickand multiresolution features,'' IEEE Access, vol. 8,

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 3, March 2020

pp. 57497–57504,2020.
[7] X. Zeng, H. Chen, Y. Luo, and W. Ye, ''Automated diabetic retinopathydetection basedon binocular siamese-like convolutional neural network,''IEEE Access, vol. 7, pp. 30744–30753, 2019.
[8] M. Manjramkar, ''Survey of diabetic retinopathy screening methods,''in Proc. 2nd Int. Conf. Trends Electron. Informat. (ICOEI), May 2018,pp. 1–6.
[9] 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 fordiabetic retinopathy detection,'' IEEE Access, vol. 7, pp. 150530–150539,2019.
[10] 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.
[11] M. Ghazal, S. S. Ali, A. H. Mahmoud, A. M. Shalaby, andA. El-Baz, ''Accurate detection of non-proliferative diabetic retinopathyin optical coherence tomography images using convolutional neuralnetworks,'' IEEE Access, vol. 8, pp. 34387–34397, 2020.
[12] H. Safi, S. Safi, A. Hafezi-Moghadam, and H. Ahmadieh, ''Early detectionofdiabetic retinopathy,'' Surv. Ophthalmol., vol. 63, no. 5, pp. 601–608,2018.pp. 232–239, May 2016