2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)

2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) | 978-1-7281-8529-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCIS51004.2021.9397115

Machine Learning Approach for Detection of Diabetic Retinopathy with Improved Pre-Processing

Ayushi Sharma*
Department of Information Technology,
Ramrao Adik Institute of Technology, Nerul
sharma.ayushinks@gmail.com

Swapnil Shinde
Department of Information Technology,
Ramrao Adik Institute of Technology, Nerul
swapnil.shinde@rait.ac.in

Imran Ismail Shaikh
Department of Information Technology,
Ramrao Adik Institute of Technology, Nerul
ishaikh422.is@gmail.com

Madhav Vyas
Department of Information Technology,
Ramrao Adik Institute of Technology, Nerul
madhav.vyas@rait.ac.in

Soumya Rani
Department of Information Technology,
Ramrao Adik Institute of Technology, Nerul
soumyarani4021998@gmail.com

***Abstract*— Diabetic Retinopathy has been one of the common reason why diabetic patients lose sight. It is caused by the damaging of the blood vessel of the light sensitive tissue at the back of the eye or the retina. The traditional methods for the detection of the disease requires lots of time and efforts hence the proposed system aims to detect diabetic retinopathy using Image Processing and Machine Learning. The input provided to the proposed system are the retinal images from the standard databases DIARETDB0 and DIARETDB1. The system aims to automate the process of detection using a combination of basic image processing steps with more focus on pre-processing to obtain clear image for feature extraction, further Machine Learning algorithms are applied for classification. The statistical features such as area and perimeter are obtained from blood vessels and exudates after pre-processing stage, different pre-processing techniques such as grayscale conversion, binarization, canny edge detection and some morphological operations including image dilation and erosion are applied to obtain clear image. The Machine Learning algorithms that are used for result analysis are Weighted KNN, Cubic SVM and Simple Tree which provides an accuracy of 85.8%, 87.2% and 88.6% respectively. The sensitivity and specificity parameters are also obtained, Simple tree yields better results for all the three result parameters.**

Keywords— Blood Vessels; Diabetic Retinopathy; Exudates; Machine Learning; Weighted KNN

I. Introduction

A. Diabetic Retinopathy

Diabetic Retinopathy (DR) is a medical condition in which a patient suffering from diabetes have some complexities in the retina, and due to these complexities, the patient's eyes are affected. Diabetic Retinopathy is caused by the damage to the blood vessels present in the retina due to diabetes. The blood vessels in the retina either swell up or contract which leads to formation of lumps of fluid or blood which are called as exudates, hemorrhage and micro aneurysm. These are the early symptoms that can be seen in the retinal images of a diabetic patient. It can also result in permanent loss of vision [2]. It is predicted that the number of diabetic retinopathy patients would increase from 7.7 million to 11 million by 2030 and to 14.5 million by 2050. A survey conducted by Rajendra Prasad Centre for Ophthalmic Sciences and AIIMS, New Delhi in 2019 shows that one out of 46 diabetic patients suffers from diabetic retinopathy. The retinal image of normal eye and DR affected eye is shown in Fig. 1.

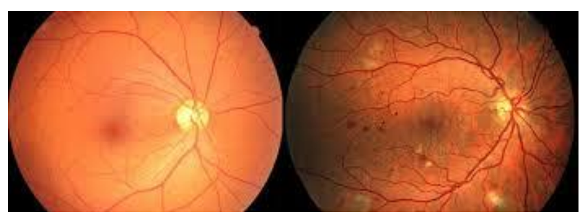


Fig. 1. Normal V/s DR affected Retina [2]

The treatment for mild cases could be done by managing sugar level of the blood and in advance cases the help of surgery or laser treatment is taken [3]. The techniques used in existing system includes Binarization, Thresholding, Opening, Adaptive Histogram Equalization, and Median Filter.

B. Problem Definition

Diabetic Retinopathy is a growing issue and the leading cause of blindness in the working age adults that is people whose age ranges between 20 and 74. Currently that statistics show that nearly 7.7 million people have suffered from diabetic retinopathy and it is predicted that the number would increase

ISBN: 978-1-7281-8529-3/21/$31.00 ©2021 IEEE

to 11 million by the year 2030 (nearly double in the next ten years) and 14.5 million by 2050. In India, the severity of diabetic retinopathy is so much that around 18% of the Indian citizen are currently suffering from diabetic retinopathy.[4] There are important differences over the past few decades in diagnosis, medical care, socioeconomic factors and other risk factors that influence the prevalence and geographic distribution of diabetes and retinopathy as well.

In statistics of the year 2002 it was witnessed that nearly 5 million people suffered from eye diseases, reason being Diabetic retinopathy which rounds up to almost 5% of world blindness. The cases of Diabetic retinopathy are gradually increasing, and hence there is a chance that probably more individual will suffer from complications in the eye, and if these complications are not properly managed, they may lead to permanent loss of vision.

The project aims to automate the process of identifying diabetic retinopathy so that preventive measures can be taken in time, and the effect of diabetic retinopathy can be minimized. Because of the automation of the process, identifying diabetic retinopathy in the early stages becomes easier and as a result, proper treatment can be provided at the proper time by the doctors.

C. Scope of the project

There are many approaches being pursued to learn more about the development and progression of diabetes-related sight loss and to identify improvements in diagnostics and treatments. Across the world, clinical trials are actively testing new options and evaluating new practices to improve the way by which diabetes-related sight loss is detected and treated. Some examples of these studies are included below.

Researchers are aware of the importance of early detection of the complications of diabetes to limit significant damage to the eye and loss of sight. As such, research teams are actively seeking improvements in the ways by which diabetes-related sight loss is identified and when it is detected. One such clinical trial is aiming to develop and test a new predictive technology that could improve the way in which Diabetic Retinopathy is identified. Researchers aim to analyze details about the individual's background, medical history and laboratory test results to develop an improved means of identifying and managing the condition.

Another study also seeks to test artificial intelligence to detect diabetic retinopathy. Research also underway seeks to find new molecules circulating in the blood stream which could highlight the prognosis of people living with diabetic retinopathy; namely identifying those people for whom the condition may be mild, moderate or more severe overtime.

The aim of the project is to identify the severity based on automated way for extraction of different features (including Blood Vessels and Exudates) of the retinal images of a diabetic retinopathy patient so that preventive measures can be taken in the earlier stage of the diabetic retinopathy.

D.Objective and Aim

- To automate the process of diagnosis of Diabetic Retinopathy
- To diagnose if the person suffers from diabetic retinopathy or not.
- To increase the efficiency of the system.

The main motive of the proposed system is to automate the whole process of detecting diabetic retinopathy .The input to the system will be the retinal images. The system will apply filtering techniques to those images. Image pre-processing steps are applied to get accurate result. Customized algorithms for the extraction of blood vessel and exudates are applied on the images. Finally the system will detect diabetic retinopathy i.e. whether the patient is suffering from diabetic retinopathy or not.

II. Literature Survey

The diabetic retinopathy concerns not only the technical aspects such as image processing and machine learning algorithms, but also the non-technical aspects such as what features to extract, and so on.

Hemanth D. et.al. [6], have proposed a hybrid method which combines image processing and deep learning on DR fundus images for improved detection. The authors have performed Histogram equalization and CLAHE on the DR images and passed it to CNN for classification. CNN provides higher accuracy but involves computation time so the authors suggest to use some pre-processing methods in image processing to reduce the same.

R. Ravindraiah et al [8] used fuzzy based clustering method, which are less complex and more robust in their system, it groups pixels with common features and accumulates data into distinct datasets. The processing steps involved in their paper included fixing the number of clusters 'C' and randomly initializing the cluster centers Ci, updating the membership matrix Mij, updating the cluster centers Ci, updating the membership matrix Mij, updating the Objective function JMKSFCM, If the absolute value of (J, t-J, t-1) is less than epsilon then stop the iteration process otherwise repeating the steps 2 to 4 Where $\epsilon = 0.0001$.

Another paper by Shailesh Kumar et al [7] proposed a system that extracts the micro aneurysm i.e. the tiny red spots on the retina for the early detection of diabetic retinopathy. The process included basic steps such as the green channel technique, adaptive histogram equalization, segmentation and SVM for classification.

The system proposed by Karkhanis Apurva Anant, Tushar Ghorpade, Vimla Jethani [9] used the methods to classify DR images using image processing, data mining approach, Algorithms such as Laplace Edge Detection, DWT and GLCM were used for the further process of feature extraction. KNN was used as the classifier which gave an accuracy of 96.62%.

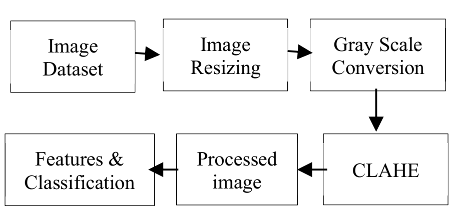
Neelam D. Panse, Tushar Ghorpade and Vimla Jethani[10] proposed a system for classifying the images into three categories normal, diabetic retinopathy and glaucoma using Discrete Wavelet Transform, Canny edge detection, Binarization, DCT (discrete cosine transformation) and KNN. The Authors [10] proposed a new method over the previously developed matched filter. The proposed system classifies the images into normal, moderate Non-Proliferative Diabetic Retinopathy (NRDR) and severe NPDR using perception based binarization, and density analysis apart from the basic image processing steps. A non-parametric Tree-type classifier-Random Forest (RF) was used as a classifier.

The system proposed by M.C. Shanker et.al. [11] discusses the detection of Diabetic Retinopathy by using SVM classifier algorithm. Pre-processing steps used in the presented system are resizing. Then the immaterial part of the images are cropped on the regions where some spots or the layer is visible. Next step was Contrast Enhancement for improving quality of the images which is necessary for human interpretation. Image Segmentation, Feature Extraction and SVM classifier were used as the steps in proposed system.
After studying existing systems it is noted that most of the systems have focused on any one of the Image processing or Machine Learning aspect of solving the problem. The proposed system tries to focus on both the processing of the image and the machine learning algorithm that can be used to detect diabetic retinopathy with much accuracy. Also, as compared to existing systems different combinations of preprocessing steps were used for each feature, blood vessel and exudates, for better extraction of the feature.

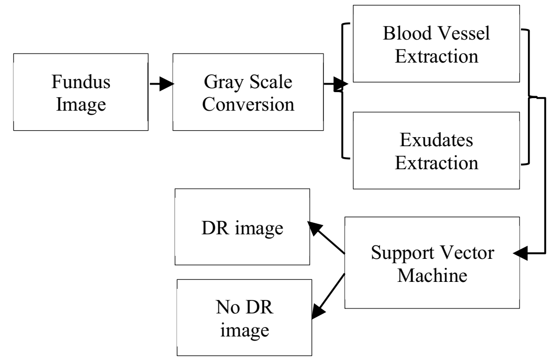
III. PROPOSED SYSTEM

The Figure 2(A) shows the proposed system diagram consisting of steps in generalize manner whereas Figure 2(B) indicates techniques used.

- **Image Resizing or Scaling:**
  Image resizing is necessary since the data set might be from different sources such as from online or directly from the hospital therefore it is necessary to bring uniformity to the images so that the processing techniques and the feature extraction technique does not result in any anomalies.



A) DR Detection Block Diagram



B) Proposed System Methodology

Fig. 2. DR Detection proposed Block Diagram & Methodology

- **Conversion of the Scale:**
  The RGB image in this step is converted to grey scale image. This is done to not only reduce the complexity while applying processing and feature extraction and other mathematical techniques if needed but also because images in grey scale the results are yielded better and a little more optimized than the former.

- **Image Pre-processing:**
  Using Adaptive histogram equalization, the contrast of the images is changed so that it will be easier to extract features.

- **Adaptive Histogram Equalisation:**
  In this technique for the betterment of the results the contrast of the image is increased. AHE also over amplifies noise in certain parts of an image. That is why an advanced algorithm is used contrast limited adaptive histogram equalization (CLAHE) which reduces the amplification of noise.

- **Feature Extraction:**
  The features of the contrasted retinal images are extracted from the above methods which would further help in classification process.

- **Classification:**
  Weighted k-Nearest Neighbor (KNN): The algorithm is customized form of k-nearest neighbor. In this algorithm a kernel function is used to give weight [12] to the nearest k points. The principle is to assign the nearby points more weight and farther points less weight.

  Support Vector Machine (SVM): The algorithm divides the training data using hyper planes and then classifies the test data using these hyper planes.

  Simple Tree: A simple tree also known as decision tree in which nodes are connected like a flowchart. The internal nodes represent an attribute and the outcomes are represented as branches.

The Algorithm steps for the proposed system for Blood vessel and Exudates detection is stated below*:*

Step 1. Input image from the user

Step 2. Apply preprocessing steps on the image for the extraction of blood vessel.

Step 2.1. Convert the image to green channel.

Step 2.2. Compliment the resultant image.

Step 2.3. Adjust the contrast of the image by histogram equalization.

Step 2.4. Open the image.

Step 2.5. Remove the optic disc.

Step 2.6. Perform Thresholding and Binarization on the image.

Step 2.7. Perform opening on the image.

Step 3. Calculate and store the area and the perimeter of the blood vessel extracted.

Step 4. Apply preprocessing steps on the image for the extraction of exudates.

Step 4.1. Convert the image to green channel.

Step 4.2. Perform Morphological closing on the image.

Step 4.3. Column Filter the image.

Step 4.4. Perform dilation then erosion on the image.

Step 4.5. Remove the optic disc.

Step 5. Calculate the area and perimeter of the exudates extracted and append the values to the above stored result.

Step 6. Input the resultant data to the pre-trained algorithm cubic SVM, weighted KNN or Simple Tree (as chosen by the user) as a test data.

Step 7. Output the data retrieved from the machine learning algorithm to the user.

IV. Implementation Output

The proposed system uses MATLAB for image processing which will be used to feed images and extract various features from the images.

A. Database

The database that is used for the system to train and test it, is DIARETDB0 and DIARETDB1. These images were captured in the Kuopio University Hospital. The database consists of 130 and 89 fundus images from DIARETDB0 and DIARETDB1 respectively out of which some images are of normal patients that does not show any symptoms of diabetic retinopathy and the rest are of the patients diagnosed with diabetic retinopathy. These images are trained and tested using MATLABR2019a.

B. Blood Vessel Extraction

Blood vessel plays an important role in determining whether the patient is suffering from diabetic retinopathy or not. Diabetic Retinopathy cause the contraction of blood vessel, but just vessel contraction is not alone capable of determining whether the patient is affected or not. This is why other features like exudate, microaneurysms are also used to detect diabetic retinopathy.

For determining whether the blood vessel has contracted or not the first step would be to extract the blood vessel from the given images. For blood vessel extraction, the steps which are used are described in the following figure, refer figure 3(A).

1) Blood Vessel extraction Process:

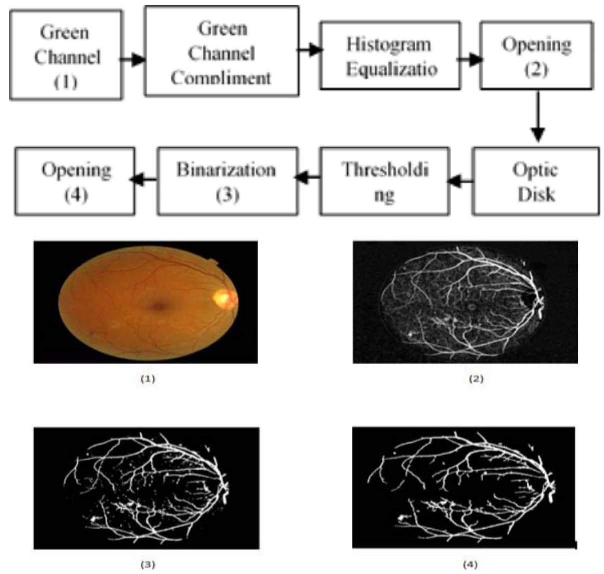


Fig. 3. Blood Vessels Extraction.

The blood vessels are extracted from the retinal image for feature extraction. Refer figure 3(B) for the results. The values obtained after calculating the area and the perimeter features are written to a CSV file.

C. Exudates Extraction

Figure 4 below shows the steps for exudates extraction. The detail explanation of each step is described below with their output shown in figure 5.

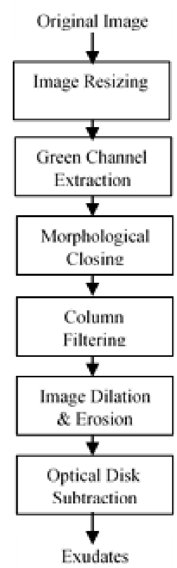


Fig. 4. Exudates Extraction- Steps.

Resize: The original Image in the RGB plane is resized to a matrix of 376 X 520 pixels, so that all the images are of a standard common size.

Green component extraction: Green Component is extracted from the RGB image for better results.

Morphological Closing: Dilation and erosion is performed on the image using a structuring element of radius 5 pixels. Refer figure 5(3).

Column filtering: The image is processed by rearranging each 5 X 5 block of the image into a temporary matrix and then applying variance (@var) function on these blocks, using this, the background reduction becomes easier. Refer figure 5(4).

Image dilation and erosion: First dilation then erosion is performed on the image to join two nearby exudates into a single bigger exudate. Refer Figure 5(5).

Optical disk masking and subtraction: Optical Disk is located and subtracted from the image. Refer figure 5(6).

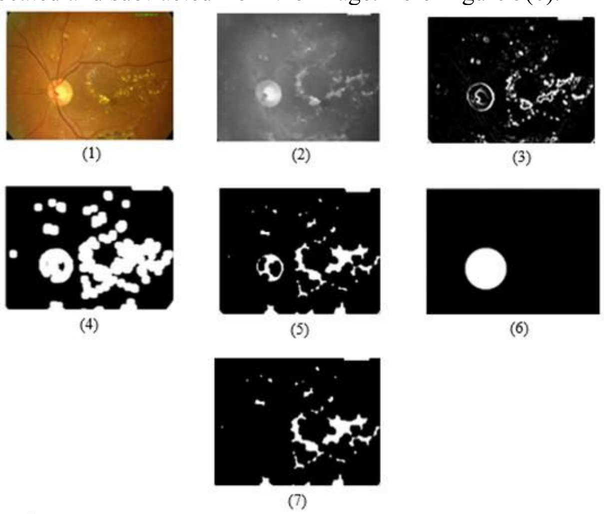


Fig. 5. Exudates Extraction – Output.

V. Experimental Results

The results that are obtained by passing the csv file obtained from the feature extraction step to the Classification Learner App that is inbuilt in MATLABR2015. The classification learner app in MATLAB takes a variable as an input in the form of predictor and response for the columns present in the table and trains the model according to the classifier selected and gives the accuracy of the classification algorithm. For this program Weighted KNN, Cubic Support Vector Machine (SVM) and Simple Tree were used to check the accuracy of the feature extraction methods. Using Weighted KNN algorithm, the overall accuracy obtained is 85.8% for exudates and blood vessel feature combination. The total number of images correctly classified overall is 188 out of 219. Table 1 represent percentage of classified images into DR and Normal categories. This class depends on the following packages for its proper functioning.

Table 1. Classification Results for KNN Classification Algorithm.

Features	Correctly Classified Image (%)	Misclassified Image (%)
Exudates	89.04	10.95
Blood Vessels	85.39	14.6
Exudates and Blood Vessels	85.84	14.16

Using Cubic SVM algorithm, the overall accuracy obtained is 87.2%. The total number of images correctly classified overall are 191 out of 219.

SVM give better results when both the features are used for classifying.

Table 2. Classification Results for Cubic SVM Classification Algorithm.

Features	Correctly Classified Image (%)	Misclassified Image (%)
Exudates	77.17	25.11
Blood Vessels	83.56	16.44
Exudates and Blood Vessels	87.21	12.79

Using Simple Tree algorithm, the overall accuracy obtained is 88.6%. The total number of images correctly classified overall are 194 out of 219.

Simple Tree too, yields better results when both the features that is blood vessel and exudates are used to classify the images, although the difference is minimal.

Table 3. Classification Result using Simple Tree Classification Algorithm.

Features	Correctly Classified Image (%)	Misclassified Image (%)
Exudates	88.13	11.87
Blood Vessels	88.13	11.87
Exudates and Blood Vessels	88.58	11.42

The following table 4 shows the accuracy results obtained using SVM, KNN and Simple Tree classification Algorithms for the proposed system.

Table 4. ACCURACY FOR MACHINE ALGORITHMS FOR PROPOSED SYSTEM

Accuracy		Classification results for feature extracted		
		Exudates	Blood Vessels	Exudates and Blood Vessels
Classifiers	Cubic SVM	77.2%	83.6%	87.2%
	Weighted KNN	89.0%	85.4%	85.8%
	Simple Tree	88.1%	88.1%	88.6%

Based on the confusion matrix values for the all the three algorithms the sensitivity and specificity is calculated for combined feature set of exudates and blood vessel in table 5.

The formula for sensitivity and specificity is given below.

$$\text{Sensitivity} = TP/\ TP+FN$$

$$\text{Specificity} = TN\ /TN+FP$$

The sensitivity and specificity values show the effectiveness of proposed system. Simple tree yields sensitivity value of 98%.

TABLE 5. SENSITIVITY AND SPECIFICITY FOR EXUDATES AND BLOOD VESSEL COMBINED FEATURES.

Algorithm	Sensitivity	Specificity
SVM	96.41	0.08
KNN	95.38	0.08
Simple Tree	97.94	0.12

The following Table 6 compares the accuracy of the proposed system with the existing system [9]. The accuracy for KNN is bit low as compared to existing system but is satisfactory. The Table 7 indicates comparison of proposed system with the paper published Kamble V & Kokate R[16] where RBF algorithm is used for classification. The results clearly indicate that the proposed DR system yields higher results and shows improved classification performance.

TABLE 6. ACCURACY COMPARISON OF PROPOSED SYSTEM AND EXISTING SYSTEM FOR KNN ALGORITHM

Algorithm	Existing System with Image Mining[9]	Proposed System(ML and Improved Pre-processing)
KNN	97.75%	85.8%

TABLE 7. ACCURACY COMPARISON OF PROPOSED SYSTEM AND EXISTING SYSTEM FOR RBF ALGORITHM

System	Classification Algorithm	Accuracy
Automated DR using Radial basis function[16]	RBF	80.0%
Proposed System(ML and Improved Pre-processing)	RBF	**88.58%**

VI. CONCLUSION AND FUTURE SCOPE

Diabetic Retinopathy being the primary causes of blindness requires reliable and accurate detection which also saves time and money of the patients. The research done in detection of DR has mainly focused on texture and statistical feature extraction as can be seen from the survey. The techniques proposed so far have focused on emphasizing the pre-processing stage more to get good quality region of interest for the next steps. Some of the methods applied include filtering, Thresholding, binarization and morphological operations. The system proposed in this paper also focuses on designing different system for blood vessel extraction and exudates extraction with more emphasis on improved pre-processing and feature extraction methods. The standard publicly available dataset of DIARETDB1 and DIARETDB0 are used for testing the proposed system design. The accuracy obtained using different algorithms is quite satisfactory, the best accuracy of 88.6% is obtained using the simple tree algorithm which was better than the other classification algorithms used for detection. The other algorithms includes Support Vector Machine (SVM) and k-Nearest Neighbor (KNN) giving an accuracy of 87.2% and 85.8%. The sensitivity and specificity for the combined features are also calculated and shown in the results, these values also support the proposed system. As a future scope researchers can apply pre-trained deep learning models for achieving higher performance in detection with replication into real time AI based models.

REFERENCES

[1] K. Verma, P. Deep, A.G. Ramakrishnan, Detection and classification of diabetic retinopathy using retinal images, in: 2011 Annual IEEE India Conference, 2011, pp. 1–6. doi: 10.1109/IND- CON.2011.6139346.

[2] B. Impairments, D. Retinopathy, Diabetic retinopathy ocutech (2020). URL:https://ocutech.com/bioptics-help-individuals-with-visual-impairments/diabetic-retinopathy/.

[3] M. Dansinger, Causes and treatments of diabetic retinopathy (2019). URL:https://www.webmd.com/diabetes/retinopathy-causes-treatments.

[4] P. Nawal, Diabetic retinopathy detection | hacker noon (2020). URL: https://hackernoon.com/diabetic-retinopathy-detection-9bdceac75752.

[5] National diabetes and diabetic retinopathy survey 2019 (2020). URL: https://currentaffairs.gktoday.in/national- diabetes- and -diabetic-retinopathy-survey-2019-102019186407.html.

[6] Hemanth, D. J., Deperlioglu, O., & Kose, U. (2020). An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network. Neural Computing and Applications, 32(3), 707-721.

[7] Shailesh Kumar, Basant Kumar, "Diabetic Retinopathy detection by Extracting Area and Number of Microaneurysm from Colour Fundus Image".

[8] Ravindraiah, R., & Reddy, S. C. M. (2018). An automated exudate detection in diabetic retinopathy fundus images using multi Kernel spatial fuzzy C means clustering method. International Journal of Engineering and Technology (UAE)., 7(18), 10-14.

[9] K. S, K. B, Diabetic retinopathy detection by extracting area and number of microaneurysm from colour fundus image, 2018 5th International Conference on Signal Processing and Integrated Net- works (SPIN), 2018.

[9] K.A. Anant, T. Ghorpade, V. Jethani, Diabetic retinopathy detection through image mining for type 2 diabetes, in: 2017 International Conference on Computer Communication and Informatics (ICCCI),2017, pp. 1–5. doi: 10.1109/ICCCI.2017.8117738.

[10] N.D. Panse, T. Ghorpade, V. Jethani, Retinal fundus diseases diagnosis using image mining, in: 2015 International Conference on Computer, Communication and Control (IC4), 2015, pp. 1–4. doi: 10.1109/IC4.2015.7375721.

[11] M. Shanker, Early detection of diabetic retinopathy using fun- dus images, International Journal for Research in Applied Science and Engineering Technology 6 (2018) 1053–1060. doi: 10.22214/ijraset.2018.3167.

[12] M. Zhao, J. Chen, Improvement and comparison of weighted k nearest neighbors classifiers for model selection, Journal of soft- ware Engineering 10 (1) (2016) 109–118.

[13] T. Kauppi, V. Kalesnykiene, J.K. Kämäräinen, L. Lensu, I. Sorri, H. Uusitalo, H. Kälviäinen, J. Pietilä, Diaretdb 0 : Evaluation database and methodology for diabetic retinopathy algorithms, 2007.

[14] T. Kauppi, V. Kalesnykiene, J. Kämäräinen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, J. Pietilä, The diaretdb1 diabetic retinopathy database and evaluation protocol, in: BMVC, 2007.

[15] J. Gou, L. Du, Y. Zhang, T. Xiong, A new distance-weighted k - nearest neighbor classifier, J. Inf. Comput. Sci. 9.

[16] Kamble, V. V., & Kokate, R. D. (2020). Automated diabetic retinopathy detection using radial basis function. Procedia Computer Science, 167, 799-808.

[17] Salman Sayad, Dr. Vandana Inamdar and Sangram Kapre,"Detection of diabetic retinopathy using Image Processing and machine Learning", 2017 IJIRSET.

[18] Karan Bhatia, Shikhar Arora, Ravi Tomar, "Diagnosis of Diabetic Retoinopathy using Machine Learning Classification Algorithm",14-16 October 2016.