A Review Paper on Prediction of Diabetic Retinopathy using Data Mining Techniques

Mrs. Pooja Rathi[1], Dr. Anurag Sharma[2]
[1]*Assistant Professor, St. Vincent Pallotti College, Raipur, India*
[2]*Associate Professor, Rungta Engineering College, Raipur, India*

***Abstract*—One of the major problem diabetic patients suffers from is the Diabetic Retinopathy and blindness. Since the number of diabetes patients is continuously increasing, this results in increase in the data as well. Hence to extract the useful information and unseen knowledge, use of data mining (DM) techniques become necessary. DM plays an important role in DR as this can be beneficial for the better health of the society. There are many techniques and algorithms that help to diagnose DR in retinal fundus images. This paper reviews, classifies and compares the algorithms and techniques previously proposed in order to develop better and more effective algorithms.**

***Index Terms*—Diabetic Retinopathy, Data mining techniques, Retinal blood vessels, Exudates**

I. INTRODUCTION

One of the major problem diabetic patients suffers from is the Diabetic Retinopathy and blindness. The greatest challenge to current health care is the rapid growth of diabetes. If detected at early stage, laser therapy can be performed to prevent or delay visual loss and may be used to encourage improvement in diabetic control. At the higher stage of diabetes, it is represented by bleeding or the accumulation of fluid in the retina. However, the symptoms can only be noticed by the diabetic patients when they start suffering from some sight disorders. If the blood vessels of the retina are damaged, it results in Diabetic Retinopathy (DR). It occurs when high blood glucose, the characteristic of diabetes, has damaged the small vessels that provide oxygen and nutrients to the retina. Diabetic Retinopathy (DR) is the leading cause of vision loss in adults aged 20–74 years. From 1990–2010, DR ranked as the fifth most common cause of preventable blindness and fifth most common cause of moderate to severe visual impairment. In 2010, of an estimated 285 million people worldwide with diabetes, over one-third have signs of DR, and a third of these are afflicted with vision-threatening diabetic retinopathy (VTDR). Every year an estimated 65–70 new cases of blindness per 100 000 occurs. These estimates are expected to rise further due to the increasing prevalence of diabetes, ageing of the population and increasing of life expectancy of those with diabetes. Recent report presented that Diabetic retinopathy is responsible for 4.8% of the 37 million cases of blindness due to eye diseases throughout the world. After 15 years, about 2% of persons with diabetes will become blind, and about 10% will develop severe visual loss. After 20 years, more than 75% of patients will have some form of DR.

Patient perception of complications : A survey says that 84% of diabetic patients knew about the complications of diabetes and 73% knew about the eye complications with 41% knowing that diabetes can lead to vision loss. Indeed, blindness was the complication patients feared the most.

Vision loss was the complication feared most by the clients interviewed (Figure 1).

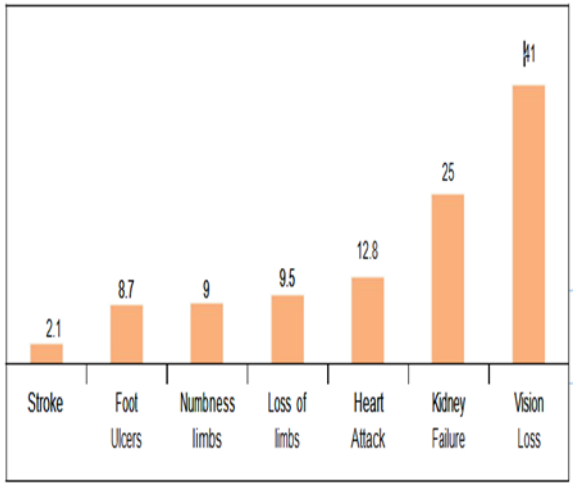

Figure 1: Complications of Concern to All Diabetics (%)

Diabetic Retinopathy is divided into four stages by Ophthalmologists. These stages are:

1. First stage is a normal retina that is free from any abnormalities.
2. The second stage is Mild (NPDR - non proliferative DR); that is, small red coloured swelling spots shown on the walls of the retina called Microaneurysm.
3. Third stage is Moderate (NPDR) where Microaneurysm begins to rupture and appears with small or medium-sized blood on the surface of the retina.
4. Fourth stage is Exudates which is called the severe (NPDR) level. exudates are of two types: Hard exudates and Soft exudates

Hard exudates have well defined boundaries. The hard exudates

I. are displayed due to the fragile and weak walls of the blood vessels of the retina. Such weak walls lead to liquid protein leakage from the blood vessels to the surface of the retina.

II. Soft exudates have unclear boundaries of a whitish colour. These exudates cause blockages in the blood vessels of the retina, and prevent the arrival of food and oxygen for its tissue. They further lead to the emergence of the so-called (Neovascularisation), winding new blood vessels;

Following Figure gives a clear picture of these four stages of DR

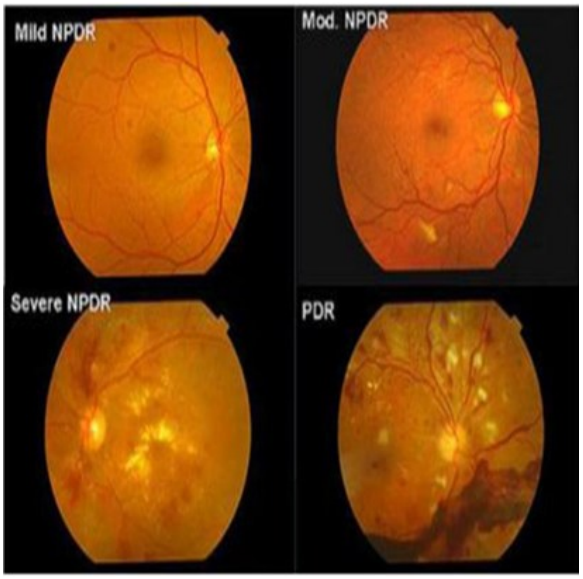

Figure 2: Diabetic Retinopathy Stages

II. ROLE OF DATA MINING

Since the number of diabetes patients is continuously increasing, this results in increase in the data as well. Hence to extract the useful information and unseen knowledge, use of data mining (DM) techniques become necessary. These techniques and algorithms are used to explore the facts and for the computations purpose. DM plays an important role in DR as this can be beneficial for the better health of the society. The focus is to discover the disease in its early stages and determine its level. In the present system, ophthalmologists have to take extra efforts and pains to diagnose such types of diabetic diseases which is time consuming too. Here comes the need for the automatic system to detect diabetic retinopathy in its early stages before it is fully blindness.

III. LITERATURE REVIEW

Dr. Karim Hashim Al-Saedi et.al.[1] have designed and implemented a system to measure the impact of diabetic retinopathy using data mining techniques In this research, an accurate measurement system of diabetic retinopathy was developed and investigated using data mining technique in which an early analysis of diabetic retinopathy using an accurate and fast technique provides the patient with enough protection treatment time. The color fundus image was used to automatically detect and realize the various lesions of diabetic retinopathy and its normal features, respectively. The specifications of the normal color fundus images were analyzed and classified by the extraction method into normal or abnormal. The abnormal image can then be categorized into three levels: Mild, moderate, and Severe. To predict the unknown class, an association rule and SVM classifier were used. The author showed promising results to support the patients and the accelerating process.

Abhilash Bhaisare et. al. [2] have proposed a system to identify the input image as normal or abnormal. When the input image is found abnormal then analysis for further DR stages is done. To identify abnormal images there are various techniques and methodology used in image mining. Image mining is an extension of data mining technique. Identification of the abnormal images is done using preprocessing, feature extraction and classification algorithms. The obtained result is used to display the image as normal or abnormal and upto what extent.

K. R. Ananthapadmanaban and G. Parthiban [3] have used Naive Bayes and Support Vercto Machine algorithms to predict early detection of eye disease and

DR. Using Rapid Miner tool they have estimated that Naive Bayes gives 83.37% accuracy and SVM gives 64.91% accuracy. Performance of these methods was also measured by specificity as 95% and sensitivity as 96.65%. They started with a preprocessing operation to improve image quality by eliminating defects caused by lighting and acquisition processes. In the second step the optic disc has disrupted the automatic detection. In the third step, the segmentation of graph cuts is used in order to detect exudates regions. Finally, the neural network gave better results with a feature extraction of images by descriptors and Hu moment of GIST. The final results were compared quantitatively with a manual exudates segmentation produced by an expert in ophthalmology.

Mahendran Gandhi et. al. [4] have performed the diagnosis of DR using morphological process and SVM classifier.Then they have applied erosion operation followed by dilation for exudates feature detection and then segmentation operation is carried out. Severe risk assessed for the degree of abnormality of an image using machine learning classifier. SVM is used to evaluate training data to find a best way to classify images into different cases like moderate or severe. In this paper, the evaluation of the automated diagnosis system of diabetic retinopathy has been performed by using a set of 5 images captured by retinal fundus cameras.

Karkhanis Apurva Anant et. al. [5] have proposed an automated eye screening system based on multi feature extraction which can detect Diabetic retinopathy based on the exudates extraction. Combination of DWT and GLCM features is used for evaluation using sensitivity, specificity parameters. This proposed system has an improved pre-processing stage which will eliminate the noise completely and thus improve overall efficiency of the system. The automated system will extract the exudates and based on the ratio of exudates to total pixels in the image will define the severity of the DR

C.Aravind et. al.[6] have proposed a new method detecting the microaneurysms from colour fundus retinal images based on feature classification. They have applied preprocessing techniques inorder to remove the optic disk and similar blood vessels using morphological operations. The preprocessed image was then used for feature extraction and these features were used for classification purpose. The classifier used is Support Vector Machine which improves sensitivity, specificity and gives an average accuracy of 90%.

Gulin Elibol and Semih Ergin [7] have presented a paper in which diagnosing the stages of DR is performed on a publicly available database (DiaraetDB1) via detecting the symptoms of this disease. Time-domain features are extracted and selected to classify a fundus image. Fisher's Linear Discriminant Analysis (FLDA), Linear Bayes Normal Classifier (LDC), Decision Tree (DT) and k-Nearest Neighbour (k-NN) are used as the classification methods in the experimental benchmarking. KNN is observed as the best classification method for without feature selection case and it gives averagely 92.22% accuracy.

Mr. Pratap Vikhe, Ms. Preeti Mistry, and Mr. Chandrakant Kadu [8] have proposed methods for detecting DR disease level emphasizing on determination of three important types of Diabetic Retinopathy; Macula Edema, Haemorrhages and Exudates. These types can be extracted using fundus images of patients and processing these fundus images through an appropriate image processing technique. Then SVM classifier is used for image classification. The method implemented in this paper can be used for screening of patient's eyeballs for detecting level of DR in a cost effective manner.

Shraddha Jalan, A. A Tayade [9], have presented a review paper on Diagnosis of Diabetic Retinopathy using KNN and SVM Algorithms. In this paper authors has mainly focused on automatic detection of Diabetic Retinopathy through detecting exudates in color fundus retinal images and also classify the lesions. they have also discussed on various methods available for detecting the exudates. Making of decision for the severity level of disease was performed by collaborating KNN and SVM classifier which gives more accuracy and reduces the time of detection of DR.

Dr. R. Geetha Ramani et. al. [10] have presented a comparative approach. This paper used two algorithms that are C4.5 Decision Tree Algorithm and Random Tree Algorithm. Both algorithms were used for comparative classification. The accuracy calculated for Decision Tree is 72.5% and for Random Tree 65%.

M. Usman Akram, Shehzad Khalid, Shoab A.Khan [11] have presented a paper showing three-stage system for early detection of Microaneurysm using filter banks. The system extracts all possible candidate regions for Microaneurysm present in retinal image. depending upon some properties, like shape, color, intensity and statistics a feature vector for each region is formed to classify a candidate region as Microaneurysm or non- Microaneurysm. A hybrid classifier which combines the Gaussian mixture model (GMM), support vector machine (SVM) and an extension of multi-model mediod based modelling approach in an ensemble is presented to improve the accuracy of classification. The true Microaneurysm regions are chosen and classified using a hybrid classifier which is a weighted combination of multivariate m-Mediods, GMM and SVM.

Vimala Balakrishnan et.al [12] have worked on Integrating association rules and case-based reasoning to predict retinopathy. they have proposed a retinopathy prediction system based on data mining, particularly association rules using Apriori algorithm, and case-based reasoning. The association rules are used to analyse patterns in the data set and to calculate retinopathy probability whereas case-based reasoning is used to retrieve similar cases. the author believed that great improvements can be provided to medical practitioners and also to diabetics with the implementation of this system.

M.Tamilarasi and Dr.K.Duraiswamy [13] have performed a survey on Automatic detection of non proliferative DR. Since there are many data mining techniques and algorithms that helps to diagnose DR in retinal fundus images, in this paper, authors have reviewed, classified and compared the algorithms and techniques previously proposed in order to develop better and more effective algorithms.

Ramon Casanova et.al.[14] focused on application of Random Forest (RF) methods to DR classification analysis based on fundus photography data. The authors proposed an approach to DR risk assessment based on metric derived from graded fundus photography and systemic data. they suggested that RF could be a valuable tool to diagnose DR diagnosis and evaluate its progression.

S.Sagar Imambi and T.Sudha [15] have applied text mining classification techniques on the available clinical data to predict the risk factor of DR. They proved that relatively short duration of case management instituted before onset of clinically identifiable retinopathy, significantly reduce the risk of developing retinopathy in patients with type 2 diabetes. The goal was to develop a scalable and robust clinical report classification system that could be applied in large hospital settings to help the physicians, so that they can guide the patients easily and reduce the vision loss.

Hayrettin Evirgen, Menduh Çerkezi [16], have used Naive Bayes classification algorithm to analyse the real life dataset in order to built predictive system for DR. A total of 385 diabetes patients' data were used to train the prediction system. With cross-validation authors proved that naive Bayes algorithm can be used for diabetic retinopathy prediction with an improved accuracy of 89%.

IV. SUMMARIZED VIEW OF VARIOUS DATA MINING TECHNIQUES

Year	Author	Technique	Dataset	Accuracy	Tool
2016	Karkhanis Apurva Anant et.al. [5]	KNN	DIARETDB1 & HRF	-	-
2016	Gulin Elibol and Semih Ergin [7]	Decision Tree, KNN, LDC	DIARAETDB1	KNN – 92.21 %, LDC – 92.45 %	MATLAB
2014	Ramon Casanova et.al. [14]	Random Forest	3443 - ACCORD Eye Study	-	WEKA
2014	K. R. Ananthapadmanaban et.al. [3]	Naive Bayes, SVM	PIMA Indian diabetic dataset	Naive Bayes – 83.3%, SVM –	Rapid miner

				64.91 %	
2014	Hayrettin Evirgen, Menduh Çerkezi [17]	Naive Bayes	Eye Clinic of the Sakarya University Educational and Research Hospital	89%	-
2013	C. Arvind et.al. [6]	SVM	Lotus Eyecare Hospital	90%	-
2012	Dr.R.Geetha Ramani et al. [10]	C4.5 Decision Tree, Random Tree	-	DT – 72.5%, Random Tree – 65%	-

V. CONCLUSION

Prediction and diagnosis of Diabetic Retinopathy depends on the presence of haemorrhages and microaneurysms in fundus images. There are many algorithms which have been proposed and developed for the automatic detection of diabetic retinopathy from feature extraction. In this paper summarized view of various data mining techniques is presented which shows that KNN and SVM have given the best accuracies. This review paper can act as a resource for the future researchers for the prediction of diabetic retinopathy using data mining techniques. This will also be useful for the researchers to get an outline of this area in order to develop more efficient algorithms.

REFERENCES

[1] Dr. Karim Hashim Al-Saedi, Dr. Razi Jabur Al-Azawi, Rasha Asaad Kamil, - Design and Implementation System to Measure the Impact of Diabetic Retinopathy Using Data Mining Techniques, International Journal of Innovative Research in Electronics and Communications (IJIREC) Volume 4, Issue 1, 2017, PP 1-6

[2] Abhilash Bhaisare, Sagar Lachure, Amol Bhagat, Jaykumar Lachure - Diabetic Retinopathy Diagnosis Using Image Mining, International Research Journal of Engineering and Technology (IRJET), Volume: 03, Issue: 10, Oct -2016

[3] K. R. Ananthapadmanaban and G. Parthiban. - Prediction of Chances - Diabetic Retinopathy using Data Mining Classification Techniques. Indian Journal of Science and Technology, Vol 7(10), 1498–1503, October 2014

[4] MahendranGandhi and Dr. R. Dhanasekaran - Diagnosis of Diabetic Retinopathy Using Morphological Process and SVM Classifier, International conference on Communication and Signal Processing, April 3-5, 2013, India.

[5] Karkhanis Apurva Anant , Tushar Ghorpade, Vimla Jethani - Diabetic Retinopathy Analysis using Image Mining to Detect Type 2 Diabetes, International Journal of Computer & Mathematical Sciences IJCMS ISSN 2347 – 8527 Volume 5, Issue 1 January 2016

[6] C. Arvind, M. Ponnibala, S. Vijayachitra - Automatic Detection of Microaneurysms and Classification of Diabetic Retinopathy Images using SVM Technique, International Journal of Computer Applications (0975–8887) International Conference on Innovations In Intelligent Instrumentation, Optimization And Signal Processing "ICIIIOSP-2013"

[7] Gulin Elibol and Semih Ergin-The Assessment of Time-Domain Features for Detecting Symptoms of Diabetic Retinopathy, International Journal of Intelligent Systems and Applications in Engineering Advanced Technology and Science, ISSN:2147-67992147, September 2016

[8] Mr. Pratap Vikhe, Ms. Preeti Mistry, and Mr. Chandrakant Kadu - Automatic Detection of

Diabetic Retinopathy Level Using SVM Technique, International Journal of Innovation and Scientific Research, ISSN 2351-8014 Vol. 11 No. 1 Oct. 2014, pp. 171-180

[9] Shraddha Jalan, A. A. Tayade - Review paper on Diagnosis of Diabetic Retinopathy using KNN and SVM Algorithms, International Journal of Advance Research in Computer Science and Management Studies, Vol.3, Issue 1, January 2015

[10] Dr. R. Geetha Ramani, B. Lakshmi, Shomona Gracia Jacob - Data Mining Method of Evaluating Classifier Prediction Accuracy in Retinal Data, 2012, IEEE International Conference on Computational Intelligence and Computing Research

[11] M. Usman Akram, Shehzad Khalid, Shoab A. Khan - Identification and classification of microaneurysms for early detection of diabetic retinopathy, Pattern Recognition, Vol. 46, No. 1, 2013, pp. 107–116.

[12] Vimala Balakrishnan , Mohammad R. Shakouri and Hooman Hoodeh - Integrating association rules and case-based reasoning to predict retinopathy, Maejo Int. J. Sci. Technol. 2012, 6(03), 334-343, ISSN 1905-7873

[13] M. Tamilarasi and Dr. K. Duraiswamy -A Survey for Automatic Detection of Non-Proliferative Diabetic Retinopathy, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1, January 2014

[14] Ramon Casanova , Santiago Saldana, Emily Y. Chew, Ronald P. Danis, Craig M. Greven, Walter T. Ambrosius - Application of Random Forests Methods to Diabetic Retinopathy Classification Analyses, Published: June 18, 2014https://doi.org/10.1371/journal.pone.0098587

[15] S. Sagar Imambi and T. Sudha - Building Classification System to Predict Risk factors of Diabetic Retinopathy Using Text mining, (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, 2010, 2309-2312

[16] Hayrettin Evirgen, Menduh Çerkezi-Prediction and Diagnosis of Diabetic Retinopathy using Data Mining Technique,ijaedu.ocerintjournals.org/tojsat/issue/22642/241874, Sep, 2014

[17] Hayrettin Evirgen, Menduh Çerkezi-Prediction and Diagnosis of Diabetic Retinopathy using Data Mining Technique, Sakarya University, Department of Computer Engineering, Serdivan Sakarya/Turkey