Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)
IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4

Implementation of Diabetic Retinopathy Prediction System using Data Mining

Siddharekh S. Patil[1], Prof. Kalpana Malpe[2]

[1]M-Tech, Department of Computer Science and Engineering, Guru Nanak Institute of Engineering & Technology, Nagpur, Maharashtra, India

[2]Assistant Professor Department of Computer Science and Technology, Guru Nanak Institute of Engineering & Technology, Nagpur, Maharashtra, India.

Abstract— **Diabetic retinopathy (DR) is the most common cause of newly diagnosed blindness every year. Annual eye checking for diabetic patients are suggested in order to find and treat DR in a timely manner, since blindness from this condition is preventable with early identification. DR detection is solely based on existing patient records. Now a day's medical data growing tremendously and we need to process that data for detection. But it is time consuming hence data mining techniques helps to get rid from this issue. We use neural network (NN) and naïve bayes for classification. According to comparison results NN gives better accuracy than naïve bayes and time and memory required for NN is less as compared to naïve bayes.**

Index Terms: Diabetic Retinopathy, Neural Network. Naïve Bayes.

I. INTRODUCTION

Number of Diabetic patients deteriorate from a medical condition in the retina of the eye called as DR. There are multiple stages of DR, which in order of enhancing austerity are: mild non-proliferative diabetic retinopathy (NPDR), moderate NPDR, severe NPDR, and proliferative DR (PDR). The main source of DR is high blood sugar levels over a long period of time in the retina called as Diabetes Mellitus. On the basis of eye test, there are enable developments in the retina as DR progresses that assist clinicians in differentiating one stage from other. These consist micro-aneurysms, intra-retinal hemorrhages, retinal ischemia, venous beading, and lastly the proliferation or growth of fragile new blood vessels that can bleed on the inner surface of retina. Implemented method will detect the presence or absence of retinopathy

The main objective is to automatically classify patients having DR and not having, given patient record as dataset. For that an initial patient records processing has been done. A classification techniques are applied in which the processed record is fed into a Neural Network and naïve bayes to predict whether the patient is diabetic or not.

Computer- supported detection of DR has been explored in the past to reduce the burden on opthamologists and alleviate diagnostic irregularities between manual readers. Automated techniques to recognize microaneursyms and DR patients records have been active areas of research . The neural networks investigated the capacity to arrange patches of ordinary retina without veins, typical retinas with veins, pathologic retinas with exudates, and pathologic retinas with microaneurysms.

The section I explains the Introduction of diabetic retinopathy prediction using methods naïve bayes and neural network. Section II presents the literature review of existing systems and Section III present proposed system implementation details Section IV presents experimental analysis, results and discussion of proposed system. Section V concludes our proposed system. While at the end list of references paper are presented.

II. LITERATURE REVIEW

Dr.V.Ramesh, R.Padmini [1] utilized classification algorithms such as MLPNN, Random Forest, Bayers net, Decision stump, Naive Bayes to predict DR risk level. They compared risk factors between patients having diabetics with and without retinopathy. Gathered data are preprocessed for different classification of risk level. They found that among different classification algorithms MLP algorithm is best suited to predict the risk factor.

Conway et al. [2] explored the job of hemoglobin level in anticipating proliferative retinopathy among 426 Type 1 diabetes patients. They utilized stereo fundus photography to decide the nearness of PDR trailed by Cox relative hazards demonstrating with stepwise regression to decide the relationship of hemoglobin level with PDR. They found that higher hemoglobin level predicts the rate of PDR, however the

affiliation fluctuates by sexual orientation, being direct and positive in men and quadratic in ladies.

AkaraSopharak, Mathew N. Dailey, BunyaritUyyanonvara, Sarah Barman, Tom Williamson and Yin Aye Moe [3] have proposed algorithm for automatic exudates detection on retinal image dataset by utilizing Fuzzy 'C' Means (FCM) algorithm. First they use pre-processing for image resizing, noise removal and contrast enhancement. To find clear exudates applied canny edge detector. Lastly find the abnormalities in rentinal images by utilizing classification such as Fuzzy 'C' Means (FCM) algorithm and K-Nearest Neighbour (KNN) classifier. Features are extracted using Gray Level Co-occurrence Matrix (GLCM) technique.
Niti Guru et al. [4] proposed system that can utilize neural networks for prediction of Heart disease, Blood Pressure and Sugar. sample database of patients' records used for examination. The ANN is tested and trained with 13 different input parameters such as Age, BP, Angiography's etc. The supervised network has been suggested for diagnosis of heart diseases.

Bagchi et al. [5] broadens the utility of asymmetric delicate margin SVM by marshaling decision theoretic ideas. It was contended this could be an increasingly sensible way to deal with set the asymmetric classification limits of a soft margin SVM. The ordinarily utilized presumption of equivalent extent of misclassifications in each class premise of assigning punishment costs for margin blunders and misclassifications were preferential. Also, the impact of info information quality on the execution of SVM is systematically explored.

Dr. Karim Hashim Al-Saedi et.al.[6] implemented a system to measure the impact of DR utilizing data mining techniques. Data mining technique utilized for an early analysis of DR using an accurate and fast technique offers the patient with enough treatment time. The color fundus image was utilized to automatically detect DR. The stipulation of the normal color fundus images were evaluated and classified by the extraction method into normal or abnormal. The abnormal image can then be divided into three levels: Mild, moderate, and Severe. To predict the unknown class, an association rule and SVM classifier were utilized.

Existing techniques are not adaptable and not sufficiently sensitive and specific to detect retinal lesions for real-life screening application. Here W. Hsu, P.M.D.S Pallawala, Mong Li Lee, Kah-Guan Au Eong [7] present the role of domain knowledge in enhancing the accuracy and robustness of detection of hard exudates in retinal images.

Some times for evaluating or predicting database is not available publically or may be cost efficient. Hence TomiKauppi et al. made DIARETDB1 [8] database publicly available for benchmarking diagnosis algorithms. With the proposed database and protocol, we can able to compare different algorithms, and correspondingly, analyze their maturity for technology transfer from the research laboratories to the medical practice.

Mahendran Gandhi et. al. [9] have implement the determination of DR utilizing morphological procedure and SVM classifier. Then they have connected disintegration activity pursued by enlargement for exudates feature finding and after that segmentation task is done. Severe hazard evaluated for the level of variation from the norm of an image utilizing machine learning classifier. SVM is utilized to assess training data to locate a most ideal approach to group images into various cases like moderate or extreme.

Enrique V. Carrera et al [10] Utilized SVM classifier to automatically detect diabetic retinopathy, they utilize image dataset for prediction. The main goal of this to detect any grade of NPDR. For that, they utilized 301 images of retina, 152 with grade 0 and 149 with grade 3. And then trained a SVM classifier with all the features of these images and then tested it through a 10-fold cross-validation process.

III. SYSTEM ARCHITECTURE

A. Architecture Overview

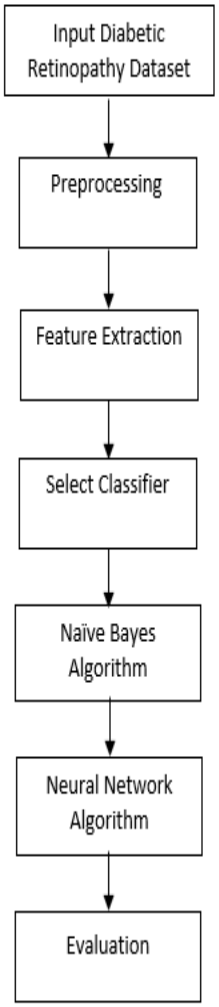

Fig 1. System Architecture

Initially we take input as a diabetic retinopathy dataset. After that preprocessing is performed on that dataset for cleaning

dataset that is removing white spaces, commas etc. Then features are extracted from processed data. Among the extracted features we select relevant features. Then we classify that data using classification technique naïve bayes and neural network for getting test output using output we can evaluate patient is predicted with DR or not.

B. Algorithms

1) Neural Network (NN)

NN is one of the learning algorithm. They consist of various layers for analyzing and learning data. Every hidden layer tries to detect patterns. When a pattern is detected the next hidden layer is activated and so on.

2) Naïve Bayes
Posterior probability:
P(R/S) =P(S/R) ×P(R) / P(S)
Where, R and S are events
P(R|S) is the posterior probability of class (R, target) given predictor (s, attributes).
P(R) is the prior probability of class.
P(S|R) is the likelihood which is the probability of predictor given class.
P(S) is the prior probability of predictor.

IV. RESULT AND DISCUSSIONS

A. Experimental Setup

All the experimental cases are implemented in Java in congestion with Netbeans tools and MySql as backend, algorithms and strategies, and the competing classification approach along with various feature extraction technique, and run in environment with System having configuration of Intel Core i5-6200U, 2.30 GHz Windows 10 (64 bit) machine with 8GB of RAM

B. Dataset Description

Diabetic retinopathy dataset downloaded from UCI machine website. Consist of patient records with symptoms detected in DR patient. The training dataset consist number of records. These records were taken from a group of subjects. The labels were given by experts who rated the presence of diabetic retinopathy in each image by a scale of 0, 1, 2, 3, 4 which indicate 0 as no DR, 1 as mild, 2 moderate, 3 severe and 4 as proliferative DR.

C. Comparison Results

This section presents the performance of the Neural network and Naïve bayes classification algorithms in terms of time or memory required and accuracy. Fig 2 Shows accuracy Comparison of naïve bayes and NN algorithms for various Threshold. X-axis shows Algorithm i.e NN, naïve bayes & Y-axis shows accuracy in %. NN gives more accurate percentage of prediction of DR than naïve bayes.

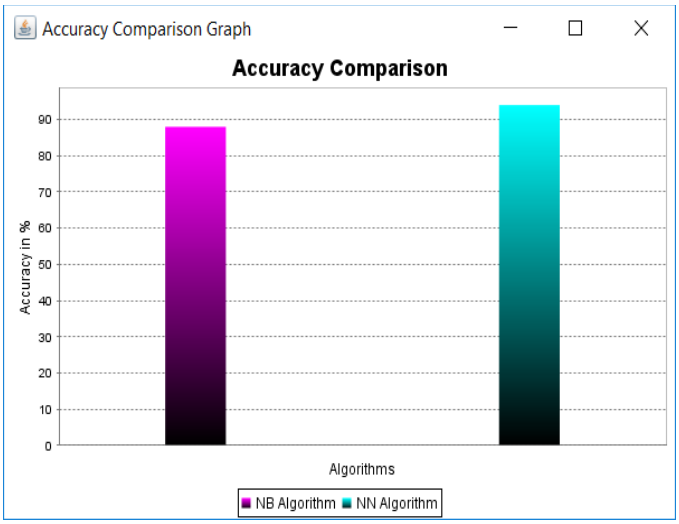

Fig. 2: Accuracy Comparison Graph

Fig 3 shows the Time comparison of naïve bayes and NN algorithms for various size. The X-axis shows algorithms NN and naïve bayes and Y- axis shows Time in milliseconds (ms). The NN takes less time than naïve bayes for classifying large text dataset.

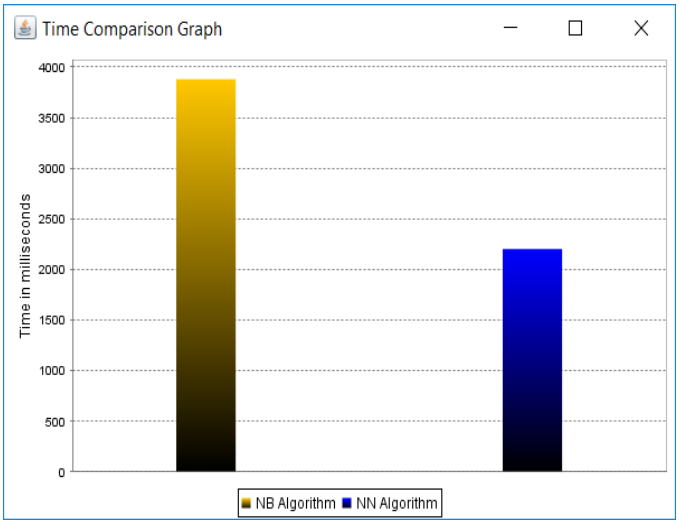

Fig. 3: Time Comparison Graph

Fig 4 indicate the Memory comparison of naïve bayes and NN algorithms. The X-axis shows algorithms NN and naïve bayes and Y- axis shows memory in bytes. The NN takes less memory than naïve bayes for classifying large dataset.

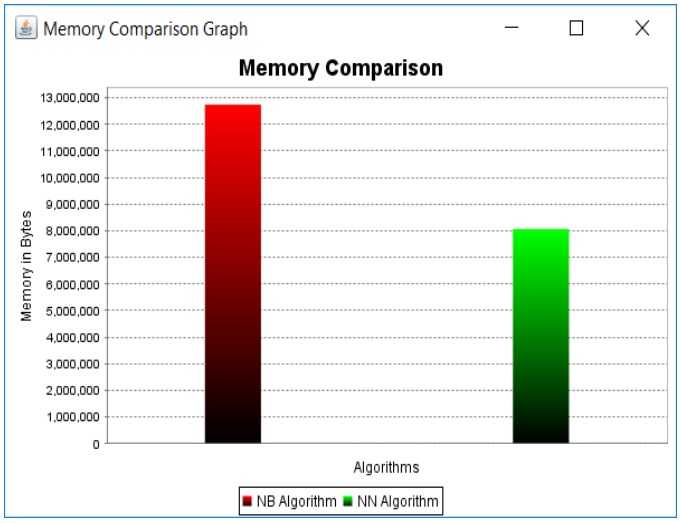

Fig. 4: Memory Comparison Graph

V. CONCLUSION

We proposed diabetic retinopathy prediction system based on some algorithms and diabetic patient sample dataset. We utilized naïve bayes and NN algorithms to classify patient data because today medical data growing and that needs to process previous records for predicting exact disease based on symptoms. We got accurate DR detected or not prediction as output, by giving the input as patients records which help us to understand the stage of DR. We compare the results between naïve bayes and NN algorithm in terms of accuracy, memory and time. The accuracy of NN algorithm which is more than Naïve bayes algorithm and time required for classification for NN is less than naïve bayes also memory required for naïve bayes is greater than the NN. Hence NN is better than naïve bayes in terms of accuracy and time.

REFERENCES

[1] Dr.V.Ramesh1, R.Padmini2, "Risk Level Prediction System of Diabetic Retinopathy Using Classification Algorithms", June 2017 IJSDR | Volume 2, Issue 6

[2] ConwayB.N., Miller R.G., and Klein R., Orchard T.J., (2009), Prediction of Proliferative Diabetic Retinopathy WithHemoglobin Level, Arch Ophthalmol, Vol. 127, pp. 1494-1499.

[3] AkaraSopharak, Mathew N. Dailey, BunyaritUyyanonvara, Sarah Barman, Tom Williamson, Yin Aye Moe, "Machine Learning approach to automatic Exudates detection in retinal images from diabetic patients", Journal of Modern optics, Vol. 57, No. 2, pp. 124-135, Nov 2015.

[4] Niti Guru, Anil Dahiya, NavinRajpal, (2007), "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. 1

[5] Tapan Bagchi, Rahul Samant, Milan Joshi (2013), ―SVM Classifiers Built Using Imperfect Training Data‖, International Conference on Mathematical Technique.

[6] 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

[7] W. Hsu, P.M.D.S Pallawala, Mong Li Lee, Kah-Guan Au Eong, The Role of Domain Knowledge in the Detection of Retinal Hard Exudates, Proc. 2013 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2, 2001, II-246 - II-251. Classification and Localization of Diabetic-Related Eye Disease by Alireza Osareh, Majid Mirmehdi, Barry Thomas, and Richard Markham.

[8] DIARETDB1 diabetic retinopathy database and evaluation protocol by Tomi Kauppi, Valentina Kalesnykiene, Joni-Kristian Kamarainen, Lasse Lensu, Iiris Sorri, Asta Raninen, Raija Voutilainen, Hannu Uusitalo, HeikkiK¨alvi¨ainen and Juhani Pietil¨a

[9] Mahendran Gandhi 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.

[10] Enrique V. Carrera, Andr´es Gonz´alez, Ricardo Carrera, "Automated detection of diabetic retinopathy using SVM", Conference Paper · August 2017 DOI: 10.1109/INTERCON.2017.8079692

[11] Huan Wang, Wynne Hsu, Kheng Guan Goh, and Mong Li Lee. An effective approach to detect lesions in color retinal images. InProc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 181–186, 2000.

[12] The electroretinogram in diabetic retinopathy by Tzekov R, Arden GB.

[13] Detection of Diabetic Retinopathy in Fundus Photographs by PavlePrentasi˘ c´.

[14] Spencer T, Olson JA, Mchrdy CK, Sharp FP, Forrester J. An image processing strategy for the segmentation and quantification of micro aneurysems in fluorescein angiograms of the ocular fundus. Comput Biomed Res. 1999; 29:284–302.

[15] Niemeijer M, Abramoff MD, van Ginneken B. Segmentation of the Optic Disc, Macula and Vascular Arch in Fundus Photographs. Medical Imaging IEEE Transactions. 2007; 26:116–127.

[16] Ege BM, Hejlesen OK, Larsen OV, Møller K, Jennings B, Kerr D. et al. Screening for diabetic retinopathy using computer based image analysis and statistical classification. Compute Method Prog Biomed.2000; 62:165–175.

[17] JayaKumari, C, & Maruthi, R. (2012). Detection of Hard Exudates in Color Fundus Images of the Human Retina. Procedia Engineering. Vol. 30, pp.297-302.

[18] Li Yafen. (2013). Automated identification of Diabetic Retinopathy Stages Using Support Vector Machine. In: Proceeding of the 32nd Chinese Control Conference 2013.

[19] Osare, A et al. (2009). A Computational Intelligence Based Approach for Detection of Exudates in Diabetic Retinopathy Images. IEEE Transactions on Information Technology in Biomedicine. 13(4), pp.535-545.

[20] Agurto, C et al. (2010). Multiscale am-fm methods for diabetic retinopathy lesion detection. IEEE Transactions on Medical Imaging. 29(2), pp.502-512.

[21] A. D. Fleming, K. A. Goatman, and J. A. Olson, "The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy," British Journal of Ophthalmology, vol 94, no. 6, pp. 706- 711, 2010.

[22] Hipwell JH, Strachan F, Olson JA, McHardy KC, Sharp PF, Forrester JV. Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool. Diabet Med 2012; 17: 588–594.

[23] Matei, Daniela, and R. Matei., "Detection of diabetic symptoms in retina images using analog algorithms," International Journal of Biological and Life Sciences, pp. 224-227, 2016.

[24] R.priya and P Aruna. Article: SVM and Neural Network based Diagnosis of Diabetic Retinopathy. *International Journal of Computer Applications* 41(1):6-12, March 2012