

# An Ensemble Model for Classification of Retinopathy Diabetic using Data Mining Technique

Amit Kumar Dewangan, Pooja Pradhan

**ABSTRACT :** Diabetic retinopathy is one type of diabetic disease which is related to eye. This is the major cause of blindness that is facing by human being. It is also one of the dangerous disease facing by human society. The diagnosis of diabetic retinopathy is very important now a days using various types of techniques. Here, there are various classification techniques and analysis tools can be to diagnosis of such type of diseases. The diagnosis of diabetes can be done using like C4.5, Random Forest, MLP, SMO, Naïve Bayes, Bayes Net and Random Tree data mining techniques for classification of Diabetic retinopathy disease. We have also ensemble the trained classifiers like: Random Forest (RF) and Multi Layer Perceptron (MLP) for achieving the better accuracy, sensitivity and specificity measures of diagnosis of diabetes retinopathy.

**Keywords :** Machine Learning, Classification, Retinopathy Diabetic, Multilayer Perceptron MLP), Random forest (RF).

## I. INTRODUCTION

Nowadays Diabetes has emerged as a major health care problem in India. According to the International Diabetes Federation (IDF), there were an estimated 40 million patients with diabetes in India in 2007 and this number is predicted to rise to almost 70 million patients by 2025. The countries with the largest number of Diabetic people will be India, China and USA by 2030 [5].

Diabetic retinopathy is the name given to the changes in the retina, which occur over a period of time in diabetics. The retina is the back part of the eye and is made up of cells, which are sensitive to light. It is fed by a network of blood vessels and any change in these blood vessels can cause the difficulties with vision. The walls of the blood vessels become fragile and then start to break, leaking blood around them. Sometimes, before the walls actually break, the weakened area can be seen, by the person who examines the eye, to have ballooned out. These are called micro-aneurysms. If these break, the amount of blood which leaks out is fairly small, and the only symptoms may be a few areas of blurring or floating spots in front of the eyes. Later the blood vessels may stop carrying blood permanently, and the cells in the retina will die from lack of nourishment. This kind of loss of sight is gradual but at the present time, it is permanent [7]. Diabetes mellitus (DM) is a multi-factorial metabolic disorder characterized by altered insulin production or activity, clinically manifested as elevated blood glucose. DM can be divided into type 1 or insulin dependent DM (IDDM) and type 2 or non-insulin dependent DM (NIDDM) [1]. Diabetic retinopathy (DR) is a serious complication of diabetes mellitus, and its prevalence has been increasing worldwide. DR remains a leading cause of preventable vision loss, despite advances in diabetes care [2].

There are various authors who have worked in the field of diagnosis of Diabetic retinopathy disease. Shahira M. Habashy et al. (2013) suggested Fuzzy C-Means (FCM) clustering for detecting and classification of Diabetic retinopathy disease [7]. Shradha Jalan et al. (2015) proposed the overall methods developed to detect exudates from retinal digital images of retinopathy patients and it is intended to help the ophthalmologists in the diabetic retinopathy screening process to detect symptoms faster and more easily [3]. Niladri Sekhar Datta et al. (2013) proposed a method to enhance the quality of the DR image while preserving the sharpness and minute details. They also proposed a method that may be directly implemented in any image processing engine both for large and high resolution Diabetic Retinopathy images [4]. Yanan, Zhang et al. (2016) compared classifiers to validate on a public diabetic retinopathy dataset. Kernel technique and bagging technique are also tested and analyzed.

## II. METHODS AND MATERIALS

This section includes data mining based classification techniques like Random forest (RF), Multilayer Perceptron (MLP) and ensemble model for classification of diabetic retinopathy disease patient or not. This section also includes description of data set used in this research work.

## 2.1 Random Forest

RF is one of the so-called ensemble methods for classification, because a committee of learners (trees in this case) is generated and each one casts a vote for the predicted label of a given instance. The trees are built using the classification and regression trees methodology (CART) [9]. In constructing the ensemble of trees, RF uses two types of randomness: first, each tree is grown using a bootstrapped version of the training data. A second level of randomness is added when growing the tree by selecting a random sample of predictors at each node to choose the best split. The number of predictors selected at each node and the number of trees in the ensemble are the two main parameters of the RF algorithm.

## 2.2 MLP

MLP (Pujari, A. K., 2001) [10] is a development from the simple perceptron in which extra hidden layers (layers additional to the input and output layers, not connected externally) are added. More than one hidden layer can be used. The network topology is constrained to be feed forward, i.e., loop-free. Generally, connections are allowed from the input layer to the first (and possibly only) hidden layer, from the first hidden layer to the second and so on, until the last hidden layer to the output layer. The presence of these layers allows an ANN to approximate a variety of non-linear functions. The actual construction of network, as well as the determination of the number of hidden layers and determination of the overall number of units, is sometimes of a trial-and-error process, determined by the nature of the problem at hand. The transfer function generally a sigmoid function.

## 2.3 Ensemble Model

Two or more models combined to form a new model is called an ensemble model (Han J. et al., 2006) [11]. An ensemble model is a combination of two or more models to avoid the drawbacks of individual models and to achieve high accuracy. In this research work, we have used various combination of Random Forest and MLP to improve the performance of model.

## III. Data Set

Diabetic Retinopathy Debrecen Data Set Data Set is a collection of UCI repository [8] which consist 19 attributes and 1 class having Diabetic Retinopathy patient and Non Diabetic Retinopathy patient. This data set also consists of 1151 instances. In which 611 are Diabetic Retinopathy patient records and 540 non Diabetic Retinopathy patient records. Class label. 1 = contains signs of DR (Accumulative label for the Messidor classes 1, 2, 3), 0 = no signs of DR.

## IV. EXPERIMENTAL RESULTS

In this research work, we have used WEKA data mining software to analysis of diabetic retinopathy data. We have compared the performance of models using WEKA data mining tools. We have used data mining based classification techniques like C4.5, Random Forest, MLP, SMO, Naïve Bayes, Bayes Net and Random Tree for classification of diabetic data but RF and MLP classifier for classification of Diabetic retinopathy. We have also ensemble these two techniques which achieved better classification accuracy compare to all individuals classifiers. The proposed ensemble classifier gives 71.1806% of accuracy which is highest among others. Table 1 shows that performance of classification techniques using WEKA data mining software and fig1 is graphical representation of result.

Table 1: Accuracy of models in WEKA data mining Tool

Algorithm	Accuracy	Confusion Matrix
Random forest	65.625 %	a b <-- classified as 91 38   a = 0 61 98   b = 1
MLP	69.7917 %	a b <-- classified as 91 38   a = 0 49 110   b = 1
RF +MLP	71.1806 %	a b <-- classified as 96 33   a = 0 50 109   b = 1

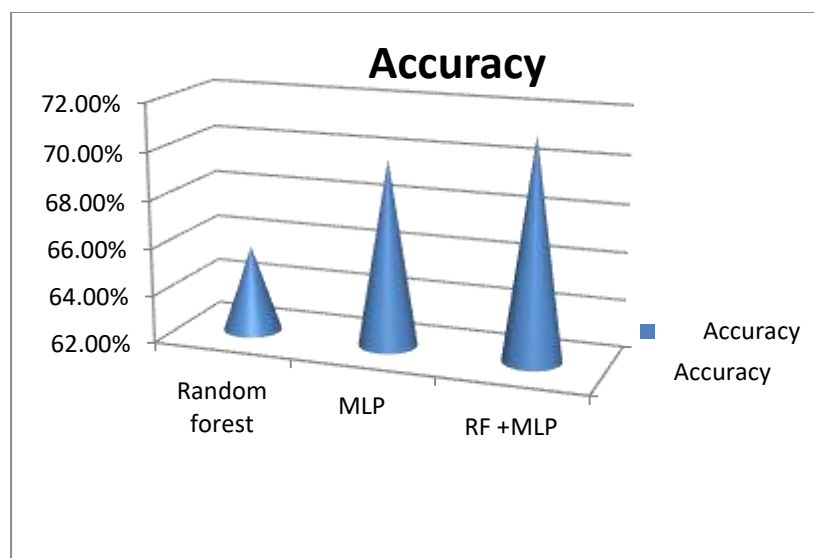


Fig. 1: Performance measures of models with WEKA data mining tools.

## V. CONCLUSION

Diagnosis of Diabetic retinopathy disease is very dangerous disease faced by most of human being. Classification is a method through which we found the Diabetic retinopathy disease patient or normal patient. In this work, we analyzed models like Bayes Net,, RF, C4.5 and Random tree gives satisfactory result. We have also developed the ensemble model by combining two techniques: Random Forest (RF) and Multi Layer Perceptron (MLP) for the improving the accuracy which gives 71.1806% of accuracy.

## REFERENCES

- [1]. Ajay Sharma, TrishnaTaye, and SB Rasel (2015) : Diabetic Retinopathy: An Overview, International Journal Of Health Research In Modern Integrated Medical Sciences, Volume 2, Issue 1, PP 36-44.
- [2]. Ting-Ting Wana, Xiu-Fen Lia, Yan-Ming Sunb, Yan-Bo Lia, and Ying Sua (2015) : Recent advances in understanding the biochemical and molecular mechanism of diabetic retinopathy. Biomedicine & Pharmacotherapy, 74, PP 145–147.
- [3]. ShraddhaJalan and A. A. Tayade (2015) :Review paper on Diagnosis of Diabetic Retinopathy using KNN and SVM Algorithms,International Journal of Advance Research inComputer Science and Management Studies, Volume 3, Issue 1,PP 128 – 131.
- [4]. NiladriSekharDatta,HimadriSekharDutta, Mallika De and SaurajeetMondal (2013) : An Effective Approach: Image Quality Enhancement for Microaneurysms Detection of Non-Dilated Retinal Fundus Image , International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) ,ScienceDirect , Procedia Technology, 10, PP 731 – 737.
- [5]. S. Wild, G. Roglic, A. Green, R. Sicree, and H. King (2004) : Global prevalence of diabetes: estimates for the year 2000 and projections for 2030 , Diabetes care. Volume 27, Number 5 , PP 1047–1053.
- [6]. Yinan, Zhang and Mingqiang An (2016) : An Active Learning Classifier for Further Reducing Diabetic Retinopathy Screening System Cost, Computational and Mathematical Methods in Medicine Volume 2016, PP 1-11.
- [7]. Shahira M. Habashy (2013): Identification of Diabetic Retinopathy Stages using Fuzzy C-means Classifier, International Journal of Computer Applications ,Volume 77 – No.9, PP 7 – 13.
- [8]. UCI Repository of Machine Learning Databases, University of California at Irvine, Department of Computer Science. Available:<http://www.ics.uci.edu/~mllearn/databases/> (Browsing date: Feb. 2017).
- [9]. Breiman L, Friedman JH, Olsen RA, Stone CJ (1984) Classification and Regression Trees: Chapman & Hall/CRC.
- [10]. A. K. Pujari, , Data Mining Techniques. Universities Press (India) Private Limited. 4th ed., ISBN: 81-7371-380-4,2001.
- [11]. J. Han and M. Kamber , Data Mining Concepts and Techniques, Morgan Kaufmann, San Francisco. 2nd ed., ISBN: 13: 978-1-55860-901-3, 2006.