iJETRM
International Journal of Engineering Technology Research & Management

A NOVEL METHOD FOR THE DETECTION OF DISEASES USING MULTI-LAYER PERCEPTRON (MLP)

Kulbir Kaur Sandhu
Assistant Professor, Department of Computer Science
Baba Farid College, Bathinda

ABSTRACT
Data Mining is a striking tool for obtaining appreciated information from huge quantity of available information that can be utilized further for taking right judgments. Numerous approaches are presented for emerging cost-effective results from the potential data. Mining of data by applying conditioning rule has predictably utilized for objective of reveling rules in medical applications. Recognition of different disease such as kidney, diabetes, and heart attack etc. from huge number of estimate and proof is an area of great attention for investigators which are not free from false assumption and unpredictable outcomes. Hence there was terrific requirement to use valuable output resulting from information of patients gathered in our data storehouse. The motive of this work is to concurrently improve parameters and feature subset without debasing classification accuracy. We introduce a genetic algorithm technique for feature selection and parameters optimization to resolve these kinds of problems. In this paper we design an algorithm that enhances accuracy of diseases prediction system named as MLP.

KEYWORDS:
Data Mining; KNN; Multi-Layer Perceptron; Accuracy; Recall; Precision; F-measure

I. INTRODUCTION

In medical domain, data mining has unlimited potential for discovering concealed patterns of data sets and further useful for clinical diagnosis. Data requires to be managed in proper manner as available data is in heterogeneous form and is in very large amount. To hide patterns in data, data mining approach provide a user-oriented technique. As said by World Health Organization, 12 million deaths take place globally each year because of Heart, Brest cancer, and kidney, diabetes diseases. Predictive analysis of diseases is considered as a significant yet complex mission that requires to be implemented precisely and professionally. The mechanization of this field would be tremendously helpful. Since all the surgeons do not keep strength in each subject and furthermore there is lack of people at some areas. Consequently, an AMD method would possibly be helpful by bringing all of them organized at one place. Correct and effective execution of automated system requires a reasonable survey of several methods and algorithms available. In this paper we mainly focus on following type of diseases: Diabetes Diseases, Heart Diseases, Kidney Diseases and Brest cancer Diseases. There are many researches had been done in this field, which we explain in the following section?

For prediction and diagnosis of diabetes one of a kind classification and clustering algorithms are utilized. On Pima India diabetes dataset,[10]SVM was applied to diagnose diabetes. Author [7] Propose that utilizing ANF Inference framework indicates a superior accuracy for diabetes diagnosis and cancer prediction. Another author [8] also gives data concerning accuracy of Naive Bayes classifier and K –means algorithm. With the help of these techniques we can achieve accuracy around 80%. Another author [11] had been advised a novel technique named Expectation–Maximization (EM0), Loyd's H way+ algorithm and genetic algorithm. EM algorithm becomes erroneous when dataset dimension is excessive [11].

For diabetes prediction, Decision tree is utilized which attain 78.17% accuracy [12]. In [12] for identification of pattern and binary classification back propagation and ANN units are utilized. In [13] prediction of diabetes had been performed with

iJETRM
International Journal of Engineering Technology Research & Management

the help of K predicting technique and Amalgam KNN- a hybrid of KNN with multistep preprocess. Fundamental statement is to strengthen K means efficiency with KNN classification for diabetes prediction. Mostly researcher give attention towards Neural Network, Data clustered algorithms, Hybrid Genetic Algorithm and Fuzzy Logic algorithm [9].
In this work, for the sake of prediction diabetic dataset is given as input; for the purpose of diagnosis fuzzy logic is utilized and at the end for accuracy evaluation hybrid genetic algorithm is utilized. In [9] author also introduces noise factors which effect neural network and accuracy in result.
In [14] analysis of diabetes is performed through Learning Machine.
Section II of the paper represent proposed methodology and last section illustrate experimental results of our proposed technique.

II. METHODOLOGY

In this research work, we basically work on sequence of dataset. For feature extraction we use Multi-Layer Perceptron (MLP).
Apart from input and output layer MLP comprises one or more hidden layers. Single layer perceptron is capable to learn linear functions, but if we want to learn non – linear functions we need multi-layer perceptron.
Below fig 1.1 indicates multi-layer perceptron with a single hidden layer. In the figure only three weights (w0, w1, w2) are indicate but in actual each connection have weights linked with them.
Input Layer: At the input layer there are three nodes. First node named bias node has fixed value i.e. 1other two nodes take external input as X1 and X2. The value of X1 and X2 are depending upon input dataset it may contain numerical values. As deliberated above, at input layer there is no computation performed, because of this the outputs from nodes in the Input layer are 1, X1 and X2 respectively, which are fed into the Hidden Layer.
Hidden Layer: Same as input layer this layer also contains three layers with initial layer Bias node having an output of 1. On the other hand output of other nodes depends upon Input layer output along with weights connected with edges. From below fig we show the output calculation for one of the hidden nodes. In the same way, other hidden layer output is computed. Reminisce that *f* denotes to activation function. These outputs are then nursed to nodes in Output layer.

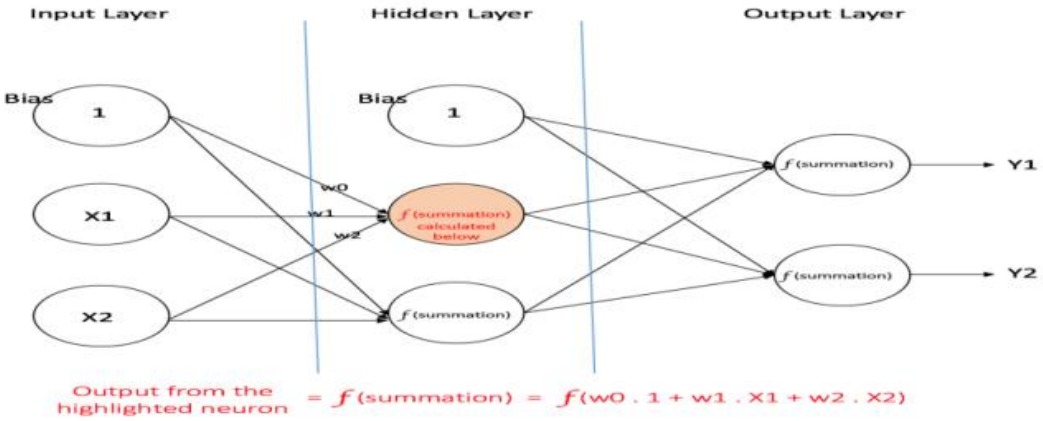


Figure 1.1: a multi- layer perceptron having one hidden layer

Output Layer: This layer contains two nodes, in which input is taken form hidden layer and perform similar computations as indicate for the painted hidden node.The values computed (Y1 and Y2) as an output of these computations performance as outputs of the Multi-Layer Perceptron.
Given a X = (x1, x2 …) features set and a target y, a Multi-Layer Perceptron can acquire relationship among features and target, for either classification or regression.

iJETRM
International Journal of Engineering Technology Research & Management

III. RESULTS

The Data Mining techniques are quite helpful in one the stages of Predictive Analysis of Diseases. In this work we mainly focus on four parameters i.e. Accuracy, Precision, Recall, F- measure.

Accuracy: This refers to the classifiers capability to properly measure intrusions from training dataset. This is explained as the ratio of correctly classified data to total classified data.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

In Which, TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

False Positive Ratio

This is one of the major factors to discover the efficiency of numerous models and also the main concern while network setup. A normal data is deliberated as abnormal or attack type data. It is demonstrated as:

$$\text{FPR} = \frac{FP}{FP+TN}$$

False Negative Ratio

This is one of the major factors utilized to define a network intrusion device's incapability to find true security events under certain circumstances. An abnormal data is not detected and considered as normal data. It is defined as:

$$\text{FNR} = \frac{FP}{FN+TN}$$

Recall: Recall is how many relevant items are selected. It is a ratio of true positive to sum of true positive and false negative. In medical diagnosis, test sensitivity (Recall) is the capability of a test to correctly identify those with the disease (true positive rate). If the test is extremely Recall and test outcome is negative you can be closely sure that they don't have disease.

$$Recall = true\ positives/\ (true\ positive + false\ negative)$$

Precision: Precision is how many selected items are relevant. It is a ratio of true positive to sum of true positive and false positive. Test specificity is ability of test to correctly recognize those without the disease (true negative rate).If the test output for an extremely specific test is positive you can be nearly sure that they really have disease.

$$Precision = true\ negatives/\ (true\ negative + false\ positives)$$

F-measure: System *F-measure* is demonstrated as weighted harmonic mean of its precision and recall, that is $F = \frac{1}{\alpha\frac{1}{P}+(1-\alpha)\frac{1}{R}}$ Where the weight $\alpha \in [0,1]$.

In our proposed method we are going to work on above four parameters. For the sake of comparison, result of our proposed classifier i.e. MLPis compared with other classifiers i.e. KNN.

iJETRM
International Journal of Engineering Technology Research & Management

Table 1.1 Comparison table

Classifiers	*Diseases / Parameters*	Kidney	Diabetes	Heart	Brest Cancer
KNN	Accuracy	93.3962	73.6912	50	89.678
	Recall	100	77.3577	50	68.127
	Precision	86.7925	66.9903	100	75.880
	F-measure	92.9293	71.8017	66.6667	61.891
MLP	Accuracy	100	100	100	100
	Recall	100	100	100	100
	Precision	100	100	100	100
	F-measure	100	100	100	100

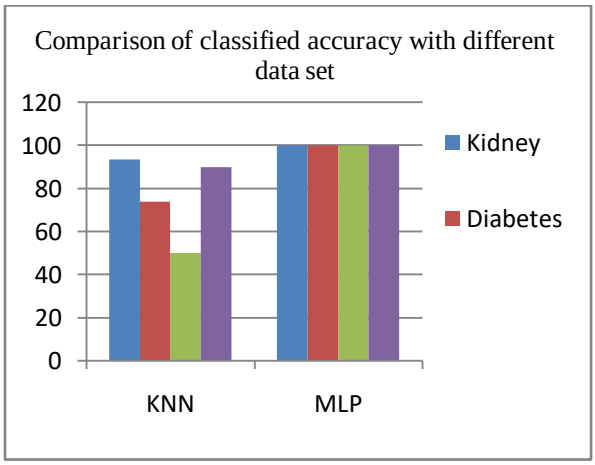


Figure 1.2: Accuracy Analysis of four Approaches

In above research, the output displays the average performance of different data sets. These models are compared on the basis of each individual fold or rounds. Experimentation result shows that the MLP approach is more accurate as associate to other existing System. Accuracy of a proposed model is 100% as shown in the Figure 1.2. This proposed method performs better existing system. To check the robustness and efficiency of any model, parameters comparison like False Positive Rate and False Negative Rate is computed.

iJETRM
International Journal of Engineering Technology Research & Management

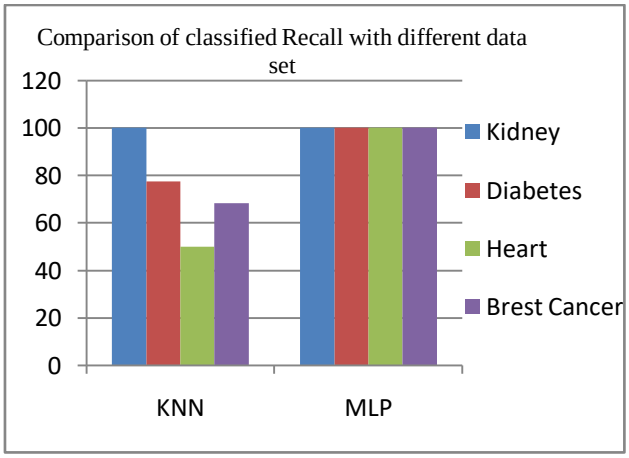


Figure 1.3: Recall Analysis of four Approaches

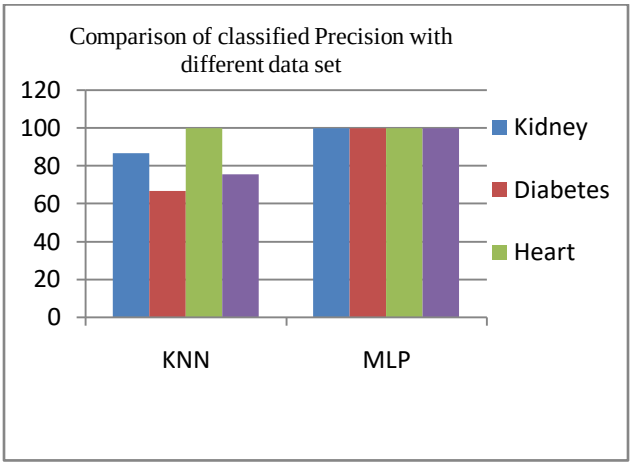
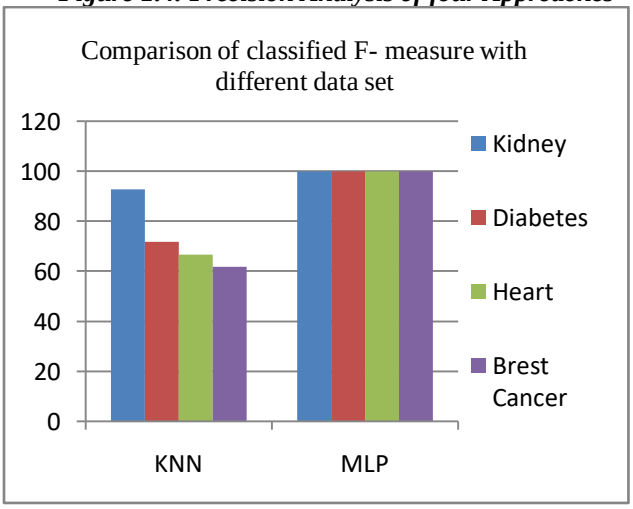


Figure 1.4: Precision Analysis of four Approaches



iJETRM
International Journal of Engineering Technology Research & Management

Figure 1.5: F- measure Analysis of four Approaches

IV. CONCLUSION

Data mining strategies had been generously utilized within field of medical diagnosis and prognosis in literature. This work notably places center of attention on feature relevance and classification systems to accurately categorize disease related to diabetes. Data mining can be used for assessing targeted disease patterns, restorative data extraction, patient backing and administration and discovering of clinical parameters. Here an algorithm is proposed that predicts four different diseases i.e. Kidney, diabetes, heart and Brest cancer which makes use of MLP. From results it is clear that our proposed, algorithm give 100% results in case of Accuracy, Precision, Recall and F- measure and improves performance of system.

REFERENCE

[1] R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines," International Journal of Medical Informatics, vol.77, pp. 81-97, 2008.
[2] Ning Wang,Guixia kang, "Monitoring system for type 2 diabetes mellitus, IEEE conference on e-health networking,pp62-67,2012.
[3] Abid sarvwar, Vinod Sharma, "Intelligent Naïve Bayes Approach to Diagnose Diabetes Type-2" Special Issue of International Journal of Computer Applications on Issues and Challenges in Networking, Intelligence and Computing Technologies,November 2012.
[4] Sonu Kumari, Archana Singh, "A data mining approach for the diagnosis of diabetes mellitus", IEEE conference on intelligent systems and control,pp-373-375,2013.
[5] Rakseh Motka,Viral Parmar, "Diabetes Mellitus Forecast Using Different Data mining Techniques IEEE International Conference on Computer and Communication Technology (ICCCT),2013.
[6] Veena Vijayan V., Aswathy Ravikumar, "Study of Data Mining algorithms for Prediction and Diagnosis of Diabetes Mellitus", International Journal of Computer Application, Vol 94,pp .12-16,June 2014.
[7] Harleen Kaur and Siri Krishan Wasan, "Empirical Study on aplications of Data Mining Techniques in Healthcare", Journal of Compuuter Science, 2006.
[8] C.kalaiselvi, G.M.Nasira, "A New Approach of Diagnosis of Diabetesand Prediction of Cancer using ANFIS", IEEE Computing andCommunicating Technologies, pp 188-190, 2014.
[9] Gunasekar Thangarasu, P.D.D. Dominic, "Prediction of HiddenKnowledge from Clinical Database using Data mining Techniques" International Conference on Computer and Information Sciences(ICCOINS),pp.1-5,2014.
[10] Velu C.M, K. R. Kashwan, "Visual Data Mining Techniques forClassification of Diabetic Patients", IEEE International AdvanceComputing Conference (IACC), pp -1070-1075,2013.
[12] Asma A. AlJarullah , "Decisiondiscovery for the diagnosis of Type IIDiabetes", IEEE conferenceon innovations in informationtechnology,pp-303-307,2011.
[13] Peter Harrington "MachineLearning in Action", ManningPublications,2013
[14] Nirmala Devi M., Appavu aliaBalamurugan S.,Swathi U.V., "Anamalgam KNN to predict Diabetes Mellitus", IEEE InternationalConference on Emerging Trendin Computing ,Communication andNanotechnology(ICECCN),pp. 691 -695,2013
[15] Jefri Junifer Pangaribuan , Suharjito "Diagnosis of Diabetes MellitusUsing Extreme Learning Machine ".pp-33-38,2014
[16] Rojalina Priyadarshini , Nilamadhab Dash , Rachita Mishra "A Novelapproach to Predict Diabetes Mellitus using Modified Extreme LearningMachine" International Conference on Electronics and CommunicationSystems (ICECS),pp.1 5,2014
[17] Sankaranarayanan.S, Pramananda Perumal T. "Diabetic prognosisthrough Data Mining Methods and Techniques" InternationalConference on Intelligent Computing Applications, pp.162-166,2014
[18] B.L. Shivakumar, Alby, S. "ASurvey on Data-Mining Technologiesfor Prediction and Diagnosis ofDiabetes", International Conference onIntelligent Computing Applications, pp.167-173,2014
[19] Poonam Undre , Harjeet Kaur , Prakash Patil "Improvement inPrediction Rate and Accuracyof Diabetic Diagnosis System UsingFuzzy Logic Hybrid Combination", International Conference onPervasive Computing (ICPC),pp. 1-4,2015.
[20] Tom Dent, "Predicting the risk of coronary heart disease", PHG foundation publisher, 2010.
[21] World Health Organization, "Global status report on no communicable diseases", 2010.