

Automatic Detection and Classification of Diabetes Using Artificial Intelligence

V. Nithyalakshmi ^a, Dr.R. Sivakumar ^b, Dr.A. Sivaramakrishnan ^c

^a *Lecturer, Jeevan Polytechnic College, Manapparai, India.*

^b *Professor, Department of Computer Science and Engineering, Tamilnadu College of Engineering, Coimbatore, India.*

^c *Associate Professor, Department of Computer Science, DMI St. Eugene University, Zambia.*

Abstract

Diabetes is characterized as a chronic disease that may cause many health complications. Artificial intelligence techniques are adopted to diagnose diabetes more accurately. This paper presents an artificial intelligence technique for diabetes diagnosis. Efficacy of the technique is evaluated using a diabetes database. Experimental results show that the back propagation neural network algorithm yields the highest classification rate compared to k-nearest neighbourhood classifier. Additionally, the back propagation neural network provides error with the highest area under curve of 90 %.

Keywords: Back Propagation Neural Network, Classification, Diabetes Diagnosis and Machine Learning Algorithm.

Introduction

The development of technology is a main root cause for decrease in physical activity. Changes in food habits and stress are also cause some health problems like blood pressure, sugar and diabetes. Diabetes is a one of the serious diseases. If cells or the pancreas do not produce adequate insulin, blood sugar increases, and it affects different organs, especially the eyes, kidneys, nerves [1] For this reason diabetes is named the silent killer.

Diabetes can be divided into three namely diabetes type I, diabetes type II and gestational diabetes. In type I, the pancreas produces very little insulin or sometimes no insulin at all. About 5 to 10% of all diabetes is type I and it occurs both childhood and adulthood. Type diabetes II occurs if the body does not produce insulin properly. Around 90% of all diabetes in the world is type II diabetes. Adults as well as children can be affected by type II diabetes. The third type of diabetes, gestational diabetes mellitus (GDM), resembles type II DM in several respects, as it involves a combination of comparatively inadequate insulin secretion and responsiveness. Around 2-10% of all pregnant women are affected by gestational diabetes, which may progress or vanish after delivery [2]. Diabetes categorization is a crucial issue for diabetes diagnosis. Medical data poses a non-linearity, non-normality and an inherent correlation characteristic. Therefore, the traditional algorithms are not suitable to classify the data properly.

Adel Al-Zebari et al. [3] compared the classification performance of various machine learning algorithms for diabetes detection. The results are evaluated on a 10-fold cross-validation basis and average classification accuracy is considered for performance measures. Pethunachiyar [4] adopted support vector machine with disparate kernel functions for diabetes classification. Kohli et al. [5] applied different machine learning models techniques for disease prediction. Feature selection is done backward modelling using the p-value test.

To increase the classification rate and generalization, Back Propagation Neural Network (BPNN) is proposed. Recently BPNN has been employed to classify disease severity via image analysis [3]. The performance has been evaluated using classification accuracy (ACC), sensitivity and specificity.

Proposed Method

1. Data Pre-processing

In this work, Diabetes dataset from the Pima Indians Diabetes Database [6] is taken for evaluation. The considered dataset is preprocessed using the image preprocessing methods and then the resulted dataset has been used for several number of experiments over different classification algorithms. The Pima Indians Diabetes Database contains the patient's details with diabetes status (Non-Diabetes and Diabetes). The considered Diabetes Mellitus dataset contains 768 records. Sample of preprocessed data is given in Table 1.

Table 1: Preprocessed Data

#Pregnant	Glucose	Blood Pressure	Body Mass Index	DPF	Age	Cases
8	183	64	23.3	0.67	32	Normal
1	89	66	28.1	0.16	21	Abnormal
0	137	40	43.1	2.28	33	Normal
5	116	74	25.6	0.20	30	Abnormal
3	78	50	31	0.24	25	Normal
2	197	70	30.5	0.15	52	Abnormal
4	110	92	37.6	0.19	32	Abnormal
10	168	74	38	0.53	35	Normal
10	139	80	27.1	1.44	49	Normal
1	189	60	30.1	0.39	56	Abnormal

After preprocessing, the size of the data is reduced to 722 records. Out of 722, 474 cases are in the class of 'Non-Diabetes,' and 248 cases are in the class of 'Diabetes' with 46 records are missing required essential values.

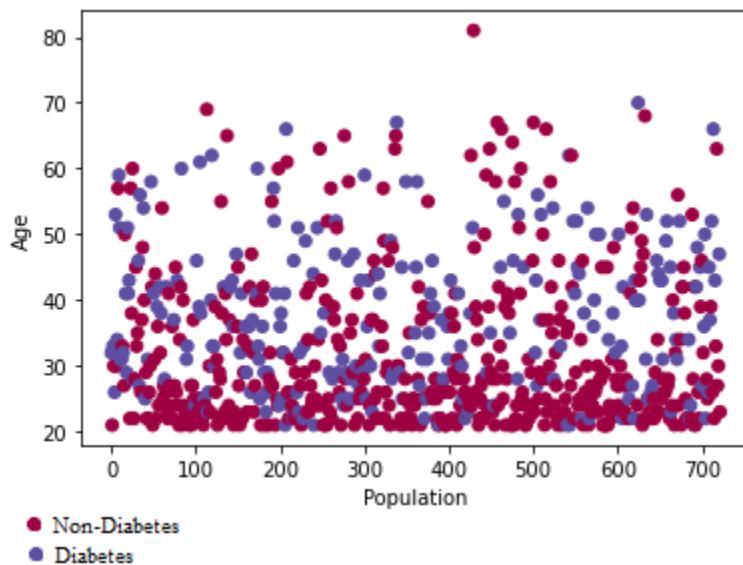


Figure 1: Age Versus Population

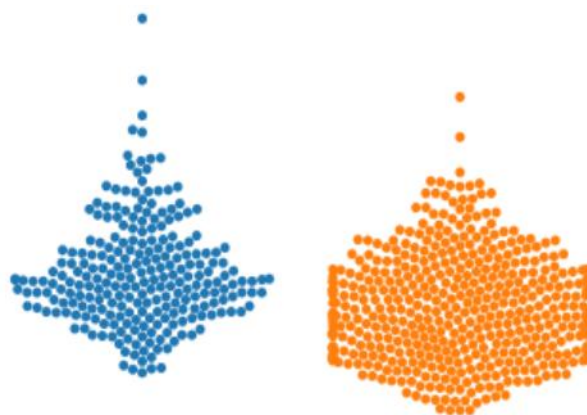


Figure 2: Body mass index Versus Diabetes

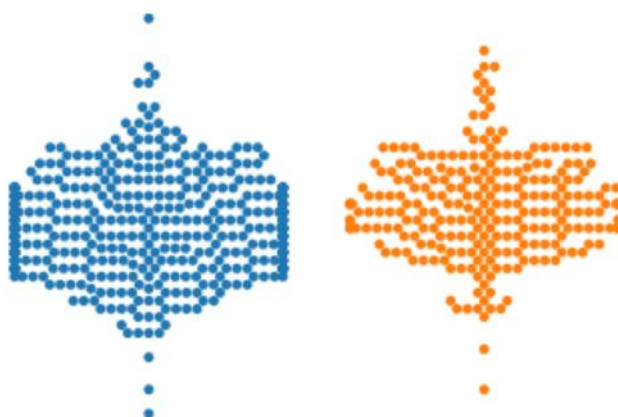


Figure 3: Blood Pressure Versus Diabetes

The patient features such as the number of times pregnant, Plasma glucose concentration, Diastolic blood pressure, Body mass index (BMI), Diabetes pedigree function (DPF), age is taken as feature for BPNN. Figure.1, Figure.2 and Figure.3 shows the population with respect to aged, diabetes/non-diabetes populations with respect to BMI and blood pressure respectively.

2. Classification

The k nearest neighbourhood (KNN) is the most straightforward, supervised machine learning algorithms used to solve classification problems. It assumes that similar things exist close. It assumes the similarity between the new data and the available data and assigns the new data to the category that is most similar to the available categories. The distance between data points is calculated using Euclidean distance; The distance between two points $(X1, Y1)$ and $(X2, Y2) = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2}$, this gives the nearest neighbourhood BPNN is a simple, commonly used neural network model, referred to as multilayer perceptron neural networks. It can be used for different due to its generalization ability.

Simulation Results

This section presents the experimental results of the BPNN and KNN. The 10-fold cross-validation is used for evaluation. In this work, the classification rate of KNN and BPNN are measured and reported. Each classification model has a different prediction accuracy based on its hyperparameters and a certain level of improvement over other prediction models. Classification outcomes are tabulated in Table.2. From the Table.3, it is observed that the BPNN yields higher prediction accuracy than KNN.

An accuracy of BPNN and KNN are 85% and 79% respectively. Similarly, Kappa value of BPNN and KNN are 0.566 and 0.234 respectively. BPNN achieved area under curve of 0.89 and KNN is 0.72. Similarly, mean square error and root mean square error of BPNN and KNN are 0.13 and 0.39, 0.27 and 0.51 respectively. Results proved that the BPNN can classify diabetes with higher accuracy and lower error. Graphical representation is provided in Figure.4 and Figure.5 respectively.

Table 2: Classification Performance of the Proposed System

Classifier	ACCURACY (%)	KAPPA	AUC	MSE	RMSE
BPNN	85	0.566	0.89	0.13	0.39
KNN	79	0.234	0.72	0.27	0.51

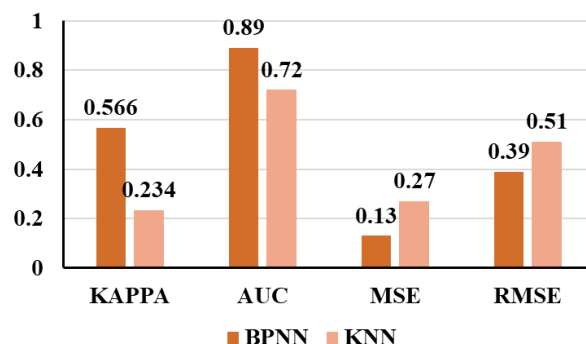


Figure 4: Efficiency of the Proposed Method

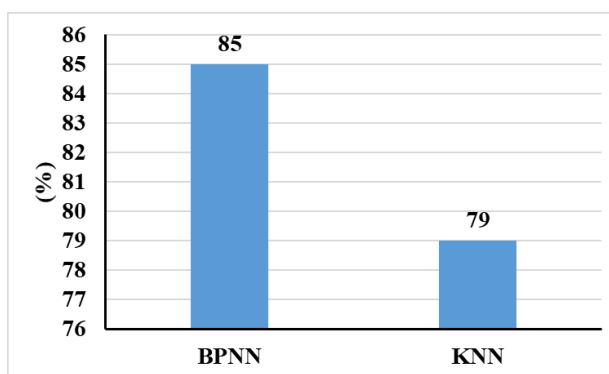


Figure 5: Classification Accuracy

Conclusion

This paper explored two machine learning models and evaluated the performances on the diabetes dataset. The results of machine learning algorithms KNN and BPNN are investigated. Several experiments are conducted and performance is measured in terms of accuracy, MSE, RMSE, Kappa score, AUROC, Results showed that BPNN has a better performance with the Prima diabetes dataset. It is suggested that the BPNN can diagnose and classify diabetic patients.

References

- Samrat Kumar Dey, Ashraf Hossain, Md. Mahbubur Rahman, "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm", *21st International Conference of Computer and Information Technology (ICCIT)*, 21-23 December, 2018.
- Usha Ruby, A., Theerthagiri, P., Jeena Jacob, I., Vamsidhar, Y., "Binary cross entropy with deep learning technique for image classification. *International Journal of Advanced Trends in Computer Science and Engineering*, 2020, 9(4), pp. 5393-5397.
- Adel Al-Zebari, Abdulkadir Sengur, "Performance Comparison of Machine Learning Techniques on Diabetes Disease Detection", *1st International Informatics and Software Engineering Conference (UBMYK)*.
- G.A. Pethunachiyar, "Classification of Diabetes Patients Using Kernel Based Support Vector Machines", *International Conference on Computer Communication and Informatics (ICCCI -2020)*, Jan. 22-24, 2020.
- Pahulpreet Singh Kohli, Shriya Arora, "Application of Machine Learning in Disease Prediction", *4th International Conference on Computing Communication and Automation (ICCCA)*.