Data Driven Analysis of Diabetes Prediction Using Machine Learning

R .Vishalakshi
Computer Science and Engineering
Velalar College of Engineering and Technology
Tindal, Erode-638012.

R .Jagapreetha
Computer Science and Engineering
Velalar College of Engineering and Technology
Tindal, Erode-638012.

R.E Aparna
Computer Science and Engineering
Velalar College of Engineering and Technology
Tindal, Erode-638012.

J.Aaron Lude Jennith
Computer Science and Engineering
Velalar College of Engineering and Technology
Tindal, Erode-638012.

Abstract-Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. It is a disease caused due to the increase level of blood glucose level. Predicting the disease at an earlier stage plays an important role to save the life of people. The primary issues we are attempting to address are increasing the accuracy of the prediction model and making the model adaptable to more than one data set. In existing system, comparison between the machine learning models versus classic auto regression in the prediction of blood glucose level is done. But as diabetes prediction. In the proposed system, we explore the seafloor popular machine eLearning models, Support Vector Machine, Decision tree, Instance based learner, multiply perceptron for the prediction of diabetes. We collected and pre-processed data then split the data into training and testing sets. We then trained the models on the training data and evaluated their performance on the data using various metrics accuracy, precision, recall and F1-score.

Keywords- MLP, IBK, SVM, J48, Preprocessing.

I.INTRODUCTION

Diabetes is one of the most common diseases in recent years, and its global prevalence is growing rapidly. It is a general term for heterogeneous disturbances of metabolism for which the main finding is chronic hyper glycaemia. The cause is either impaired insulin secretion or impaired insulin action or both. The chronic hyper glycaemia of diabetes is associated with long-term damage, dysfunction, and failure of various organs, especially the eyes, kidneys, nerves, heart, and blood vessels. The vast majority of diabetes can be divided into two categories, viz. Type 1 and Type 2. The cause of Type 1 diabetes is an absolute deficiency of insulin secretion. On the other hand, Type 2 diabetes is much more prevalent, and the cause is a combination of resistance to insulin action and an inadequate compensatory in sulin secretary response. The most common form of diabetes is Type 2 diabetes.

II.PROPOSED SYSTEM

The major goal of this proposed effort is to categorise data as diabetes or non-diabetic and increase classification accuracy. For many classification problems, the greater the number of samples picked, the worse the classification accuracy. This survey examined several categorization strategies for diabetes and non-diabetic data. As a result, it is discovered that approaches like IBK, SVM are most suited for constructing the Diabetes prediction system. With the use of new computational approaches, machine learning has the potential to change diabetes risk prediction. Diabetes must be detected early in order to be treated.It proposed a machine learning method for predicting diabetes

levels. The method may also assist researchers in developing an accurate and useful tool that will reach physicians' tables to assist them in making better decisions about illness state.

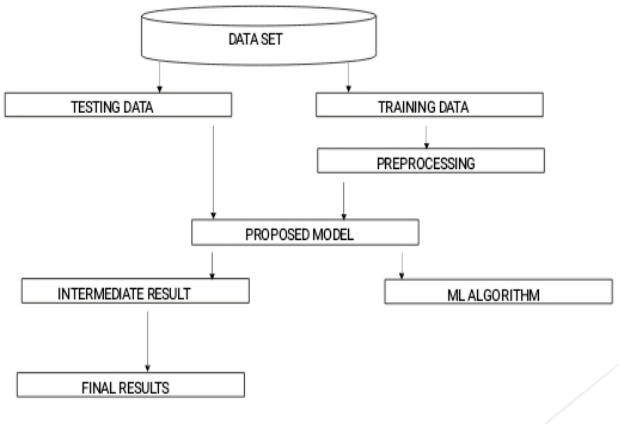

Fig.2.1 System Architecture

2.1 PREPROCESSING

It is used in the Data Selection process to detect issues such as missing numbers, incorrect content, and data inconsistency. It computes the data to achieve the desired findings in the data analysis stage by evaluating the datasets with a particular tool like WEKA.The methods that are used in this stage are imputation and cleaning.

2.2 NORMALIZING THE DATA

Identifying the incorrect values because real-world data is filthy, fragmentary, and noisy, we must use data preparation techniques. This approach entails identifying mistakes and missing data values from a large dataset. Pre-processing makes it simple to return missing values and fix erroneous data.

2.3 IMPLEMENTATION OF MACHINE LEARNING ALGORITHM

Different machine learning algorithms like SVM, J48, IBK, and MLP were applied to predict the result.

2.4 OUTPUT DATA

The output data has the information about whether a person has diabetes or not. It also gives a comparison between the algorithms used and shows which algorithm produces accurate resulting the form of bar graph.

III.ANALYSIS & COMPARISON

Accuracy, prediction, f1 Score, Recall are the factors that are compared for the machine learning algorithms (SVM, J48, MLP, IBK). It shows which algorithm is providing correct values by comparing the performance of the algorithm. By seeing the graph it is understandable that IBK is better compared to other algorithms.

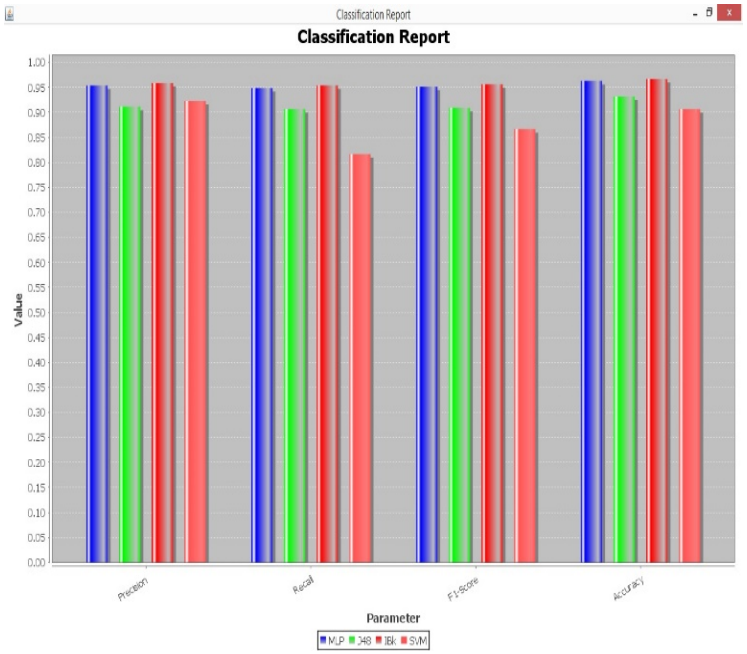

fig .3.1 Performance of algorithms

IV.EXPERIMENTAL RESULT:

ALGORITHMS	PRECISION	RECALL	ACCURACY	F1SCORE
MLP	0.9538461538461539	0.9489795918367347	0.9630350194552529	0.9514066496163682
DTJ48	0.9114583333333334	0.9067357512953368	0.9319066147859922	0.9090909090909092
IBK	0.9589743589743589	0.9540816326530612	**0.9669260700389105**	0.9565217391304348
SVM	0.9230769230769231	0.8167539267015707	0.9066147859922179	0.8666666666666667

V.CONCLUSION

The goal of this project was to design an efficient model for the prediction of diabetes. After a careful study of other published work, we proposed a novel model, which consists of using Support Vector Machine (SVM),Instance Based Learner(IBK), Dexcision tree(J48), Multilayer perceptron(MLP) for classification. By comparing all these algorithms, we conclude that IBK gives best result withanaccuracyof96.7%forthepredictionofdiabetes.

REFERENCES

[1] R. Hovorka et al., "Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT," Amer. J. Physiol.-Endocrinal. Metabolism, vol. 282, no. 5, pp. E992–E1007, 2002.
[2] C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.
[3] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM"SurajPunj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756
[4] A. Gain, A. V. Grysbok, S. Raja Raman, W. K. Ward, and J. Reitman, "Predicting subcutaneous glucose concentration in humans: Data-driven glucose modeling," IEEE Trans. Biomed. Eng., vol. 56, no. 2, pp. 246–254, Feb. 2009.
[5] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[6] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" SurajPunj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749
[7] M. Eren-Oruklu, A. Cinar, L. Quinn, and D. Smith, "Estimation of future glucose concentrations with subject-specific recursive linear models," Diabetes Technol. Therapeutics, vol. 11, no. 4, pp. 243–253, 2009.
[8] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.
[9] G.Neelakrishnan, P.Iraianbu, T.Abishek, G.Rajesh, S.Vignesh, "IOT Based Monitoring in Agricultural" International Journal of Innovative Research in Science, Engineering and Technology, March 2020, Volume 9, Issue 3, pp:814-81
[10] G. Sparacino et al., "Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series," IEEE Trans.Biomed. Eng., vol. 54, no. 5, pp. 931–937, May 2007.
[11] Dr.C.Nagarajan,G. Neelakrishnan, V.Sundarajan, and D.Vinoth, "Simplified Reactive Power Control for Single-Phase Grid-Connected Photovoltaic Inverters" International Journal of Innovative Research in Science, Engineering and Technology, May 2015; 4(6): 2098-2104
[12] K. Turksoy et al., "Adaptive multivariable closed-loop control of blood glucose concentration in patients with type 1 Diabetes," in Proc. Amer. Control Conf., 2013, pp. 2905–2910.