iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** V **Month of publication:** May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34143

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

A Review of Machine Learning Algorithms to Diagnose Diabetes Disease

Dhwani Chouksey[1], Preeti Rai[2]
[1, 2]*Department of Computer Science & Engineering, Gyan Ganga Institute of Technology and Sciences, Jabalpur*

Abstract: ***According to the WHO, approximately 422 million people worldwide suffer from diabetes, mostly live in low and middle-income countries, and diabetes causes 1.6 million deaths each year [1]. Diabetes is a metabolic disease caused by high blood sugar levels. This causes various types of disorders, for example, coronary failure, blindness, urinary organ disease, etc. It can also damage nerves, eyes, kidneys, and other organs. In such a situation, the patient is required to visit the diagnostic center, so that his report can be obtained after consultation. All the time they have to invest their time and money. With the rapid development of machine learning, it is being used in many aspects of medical diagnosis. Many researchers have developed machine learning algorithms to diagnose diabetes in the early stages. In this paper, we will review various types of machine learning algorithms and techniques.***

Keywords: Diabetes Disease, Machine Learning, Algorithms, Diagnosis, Review, Pima Indians Diabetes Dataset.

I. INTRODUCTION

In the last two decades, diabetes is one of the major emerging diseases in India. The latest study showed that diabetes increased by 45% globally and increased by 12.3% between 1990 and 2013 in India [2]. Various surveys over the last decade have shown that 10% of Indian adults suffer from diabetes and have a similar proportion of glucose tolerance. A 1991 survey showed a diabetes prevalence of 8.3% based on fasting blood glucose. The National Health Survey conducted in 2000 indicated that there was a significant increase of 11.6% in diabetes. Diabetes is a serious health problem that causes high blood sugar and the measurement of sugar content cannot be controlled. The hormone insulin transfers blood sugar to cells that are stored or used for energy. With diabetes, the body either does not make enough insulin or does not use insulin effectively. However, various factors such as sugar concentration, height, weight, hereditary factors, and insulin also play an important role in influencing diabetes.

There are three types of Diabetes:

1) *Type-1 Diabetes:* This type of diabetes is caused when the immune system attacks and destroys cells in the pancreas, where insulin is produced. It is mostly inherited and can be prevalent throughout life and therefore insulin shots or insulin pumps are used to maintain blood insulin levels. According to fact files of the World Health Organization (WHO) [3], type 1 diabetes cannot be prevented. Some common symptoms of type 1 diabetes can include extreme hunger, increased thirst, unintentional weight loss, frequent urination, blurry vision, tiredness.
2) *Type-2 Diabetes:* It is also known as 'Insulin Resistance 2'. Type 2 diabetes occurs when the body stops reacting to the insulin produced. Various lifestyle factors are responsible for the development of type 2 diabetes. According to the World Health Organization fact files, Type 2 diabetes can be prevented with a healthy diet and physical activity [3]. Some common symptoms of type 1 diabetes can include extreme hunger, increased thirst, unintentional weight loss, frequent urination, blurry vision, tiredness.
3) *Gestational Diabetes:* This is a condition that occurs when women are pregnant. This type of diabetes occurs when the insulin-blocking hormone is produced by the placenta. The condition is usually detected during a blood sugar test or oral glucose tolerance test that is usually performed between the 24th and 28th week of gestation. Gestational diabetes is completely curable with proper medication throughout pregnancy, but women who have it may develop type-2 diabetes in the future.

High blood sugar causes damage to organs and tissues in your body. The higher your blood sugar and the longer you stay with it, the higher your risk of complications. The Complications associated with diabetes include heart disease, heart attack, and stroke, neuropathy, nephropathy, retinopathy and vision loss, hearing loss, foot damage such as infections, sores that don't heal, skin bacterial, and fungal infections, depression, dementia. But with the increase in machine learning methods we have the flexibility to search for answers to the current issue, we have received advanced system mistreatment information processing, which can predict whether a patient has the disease. Also, it becomes important before starting the diagnosis that is provided to patients. Information withdrawal features the removal of large amounts of unseen data related to diabetes. The purpose of this analysis is to review the machine learning system that can estimate the patient's level of diabetes risk with better accuracy.

II. OBJECTIVE

Machine learning (ML) can help reduce inefficiencies of corrective systems ensuring a largely streamlined and cost-effective health ecosystem. The challenge is to deliver ML-based quality healthcare solutions to the general public on a large scale and make them aware of the disease. Most people do not know whether they have diabetes or not. Our aim of this study is, to analyze different types of machine learning classification models and algorithms to diagnose diabetes in the early stage.

III.LITERATURE REVIEW

In recent times, machine learning and data mining techniques have been considered to design automated diagnostic systems for diabetes. Recently, many methods and algorithms are used to mine biomedical datasets for hidden information including Neural Networks, Decision Tree, Naive Bayes, SVM, Logistic Regression, and so on. These algorithms reduce the time spent to process symptoms and produce diagnoses, making them more accurate at the same time. There are various methods related to the diagnosis and classification of diabetes disease. Some of the following research has been reviewed in the proposed system: -

TABLE I
Comparison Various Research Algorithm

Authors	Methodology	Accuracy
M. Nilashi et al.[4]	SOM, PCA, Neural Network	92.28 %
Polat et al. [5]	GDA and LS-SVM	82.05 %
Kamer Kayaer and Tulay Yildirim [6]	Meta-DermDiagnosis network	83.70 %
V. Anuja Kumari and R.Chitra [7]	Support Vector Machine (SVM)	78.00 %
Priyanka Sonar and K. Jaya Malini [8]	Decision Tree, Native Bayes, Support Vector Machine	85.00 %
Veena Vijayan V. and Anjali C.[9]	AdaBoost algorithm	80.72 %
P. Suresh Kumar and V. Umatejaswi [10]	Decision Tree, SVM, Naive Bayes	96.30 %
Minyechil Alehegn, Rahul Joshi & Dr. Preeti Mulay [11]	Ensemble Method	90.36 %

1) M. Nilashi et al.[4] proposed the hybrid intelligent system, developed through clustering, noise removal, and classification approaches in their paper "Accuracy Improvement for Diabetes Disease Classification: A Case on a Public Medical Dataset". They used SOM for clustering, PCA used for dimensionality reduction and noise removing and Neural network for the classification task. Using 10-fold cross validation they achieved an accuracy of 92.28% with the combination of SOM, PCA, and Neural Network.

2) Polat et al. [5] proposed a cascade learning system based on Generalized Discriminant Analysis (GDA) and Least Square Support Vector Machine (LS-SVM) in their research “A cascade learning system for classification of diabetes disease”. They divide the system into two stages. In the first stage, the Generalized Discriminant Analysis was used to discriminant feature variables between healthy and diabetic patient data as a pre-processing process. And in the second stage, they used Least Square Support Vector Machine for the classification task. Using 10-fold cross-validation, they achieved 82.05% of accuracy on the Pima diabetes dataset.

3) Kamer Kayaer and Tulay Yildirim [6] have proposed three different neural network architectures in their research "Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks". The first is the multilayer perceptron (MLP), the second is the radial basis function (RBF) and the third is the general regression neural network (GRNN). They applied these models to Pima Indians Diabetes (PID) medical data. They achieved the best result using GRNN structure with 80.21% accuracy on test data.

4) V. Anuja Kumari and R.Chitra [7] proposed a classification system of diagnosis of diabetes using Support Vector Machine (SVM) in their research "Classification Of Diabetes Disease Using Support Vector Machine". Using 10-fold cross-validation they achieved an accuracy of 78%, Sensitivity of 80%, and the Specificity of 76.5%.

5) Priyanka Sonar and K. Jaya Malini [8] have used different machine learning approaches for the classification model to diagnose diabetes disease in their research. They develop model-based categorization methods as Decision Tree, ANN, Naive Bayes, and

SVM algorithms. Decision Tree achieves an accuracy of 85%, Native Bayes achieve 77% and Support Vector Machine (SVM) model achieve 77.3%.

6) Veena Vijayan V. and Anjali C. [9] proposed a decision support system that uses the AdaBoost algorithm with Decision Stump as the base classifier for classification. They also implemented Support Vector Machine, Naive Bayes, and Decision Tree as base classifiers for verifying the accuracy of the AdaBoost algorithm. They achieved accuracy for AdaBoost algorithm with decision stump as the base classifier is 80.72% which is greater compared to that of Support Vector Machine, Naive Bayes, and Decision Tree.

7) P. Suresh Kumar and V. Umatejaswi [10] have proposed machine learning algorithms such as Decision Tree, SVM, Naive Bayes for identifying diabetes using data mining techniques. They cluster the dataset into 3 clusters. first for gestational diabetes, second cluster for type-1 diabetes (juvenile diabetes), and third cluster for type-2 diabetes. They achieved an accuracy of 96.3% using the Random Tree algorithm.

8) Minyechil Alehegn, Rahul Joshi & Dr. Preeti Mulay [11] discuss the algorithms like SVM, Naïve Net, DecisionStump and the ensemble method in their research "Analysis and Prediction of Diabetes Mellitus using Machine Learning Algorithm". The model achieves an accuracy of 90.36% using ensembling.

IV. FUTURE SCOPE

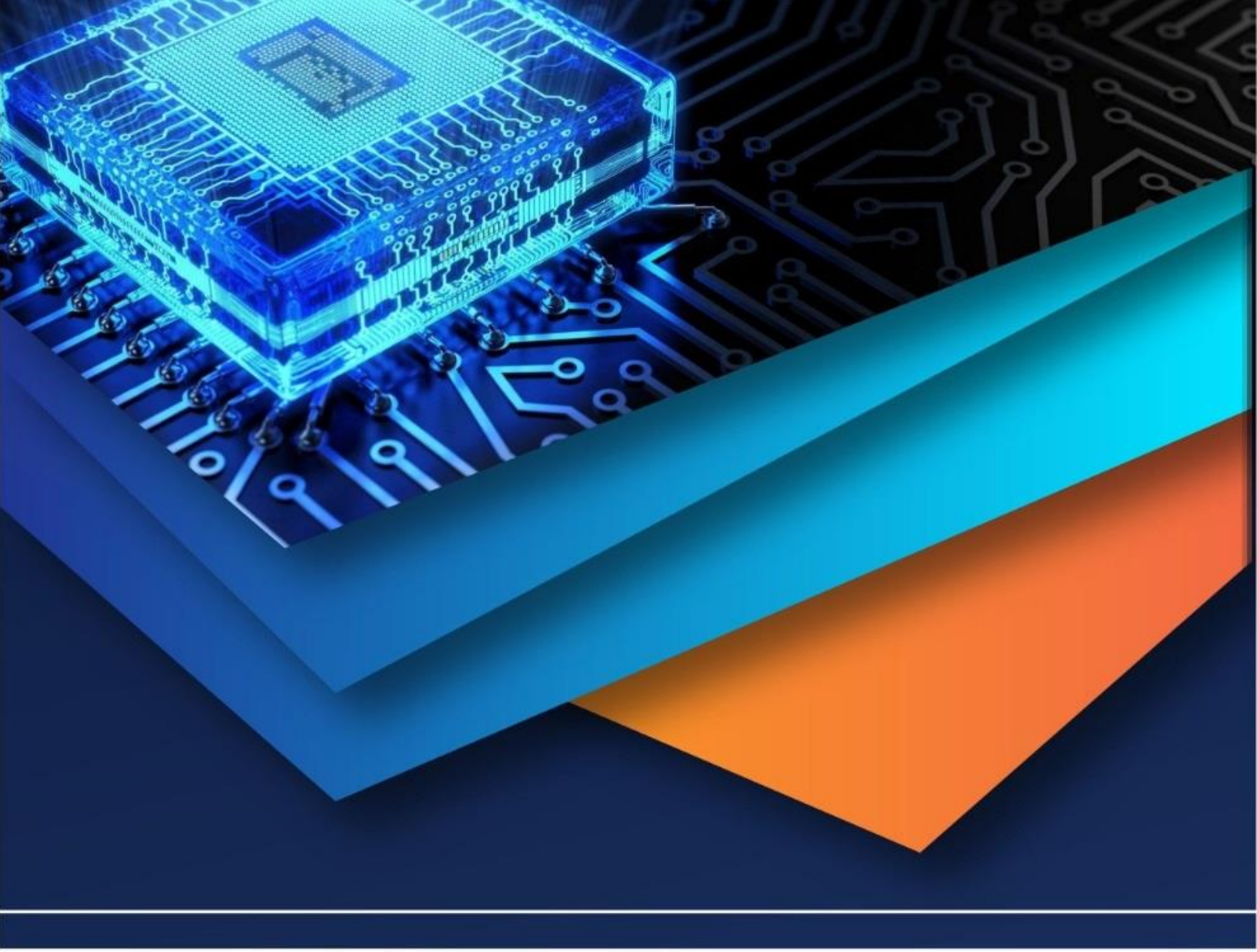
There is a lot of research done on the classification of diabetes using machine learning algorithms and techniques. In this paper, we study and analyze different types of algorithms like Support Vector Machine (SVM), AdaBoost, Neural Network, Decision Tree, Native Bayes, Ensemble Method, and so on. There is still much work to be done on the methods of clustering, noise removal, and classification for the diagnosis of diabetes disease. In future work, more attention should be paid to datasets for disease classification by using the incremental machine learning approaches. We will plan the proposed method that combines machine learning and deep learning algorithms to achieve better accuracy in data preprocessing and classification. Also, a user-friendly Graphical User Interface (GUI) application can be developed that can be used by peoples.

V. CONCLUSIONS

Detection of diabetes in its early stage is one of the important real-world problems in the medical field. It is important to make the initial diagnosis properly to prevent diabetes. we review many types of research in order to classify diabetes in the early stage. All research and machine learning models were evaluated on the Pima Indian Diabetes (PID) database. It is noticed that the accuracy of the machine learning algorithm increases with the pre-processing technique on the data. We study different types of data pre-preprocessing techniques like PCA and Clustering in the literature survey. Also observed that the ensembling technics are performed batter result in classification. It is also observed that in classification tasks, the ensembling techniques provide better results. In our future work, we combined all these technics and algorithms to achieve the best results in the diagnosis of diabetes disease.

REFERENCES

[1] https://www.who.int/health-topics/diabetes
[2] Sushmi Dey. “India’s diabetes rate up 123% since 1990, Times of India. 15 June 2015.
[3] https://www.who.int/news-room/fact-sheets/detail/diabetes
[4] Mehrbakhsh Nilashi et al., “Accuracy Improvement for Diabetes Disease Classification: A Case on a Public Medical Dataset ”, May 2017, http://dx.doi.org/10.1016/j.fiae.2017.09.006
[5] K. Polat et al., “A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine”, Expert systems with Applications 34 (2008) 482-487.
[6] Kamer Kayaer, Tulay Yildirim , “Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks”, January 2003
[7] V. Anuja Kumari and R.Chitra , “Classification Of Diabetes Disease Using Support Vector Machine”., Vol. 3, Issue 2, March -April 2013, pp.1797-1801
[8] Priyanka Sonar and K. Jaya Malini, “Diabetes Prediction Using Different Machine Learning Approaches”, Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)
[9] Veena Vijayan V. and Anjali C., “Prediction and Diagnosis of Diabetes Mellitus –A Machine Learning Approach”, Proceedings of Recent Advances in Intelligent Computational Systems (RAICS) IEEE, 2015
[10] P. Suresh Kumar and V. Umatejaswi, “Diagnosing Diabetes using Data Mining Techniques”, International Journal of Scientific and Research Publications, Volume 7, Issue 6, June 2017 705 ISSN 2250-3153.
[11] Minyechil Alehegn, Rahul Joshi & Dr. Preeti Mulay , “Analysis and Prediction of Diabetes Mellitus using Machine Learning Algorithm”, International Journal of Pure and Applied Mathematics ISSN: 1311-8080



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY