# REVIEW ON STATISTICAL DATA ANALYSIS ON DIABETES USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

**Shweta Kamble[1], Abhishek Devkhile[2], Pooja Phule[3], Prof. Shital Patil[4]**

[1-3]*B.E. Student, Dept. of Computer Engineering, Suman Ramesh Tulsiani Technical Campus Faculty of Engineering, Pune, Maharashtra, India*
[4]*Professor, Dept. of Computer Engineering, Suman Ramesh Tulsiani Technical Campus Faculty Of Engineering, Pune, Maharashtra, India*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** *Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. According to International Diabetes Federation 382 million people are living with diabetes across the whole world. By 2035, this will be doubled as 592 million. Diabetes mellitus or simply diabetes is a disease caused due to the increase level of blood glucose. Various traditional methods, based on physical and chemical tests, are available for diagnosing diabetes. However, early prediction of diabetes is quite challenging task for medical practitioners due to complex interdependence on various factors as diabetes affects human organs such as kidney, eye, heart, nerves, foot etc. Data science methods have the potential to benefit other scientific fields by shedding new light on common questions. One such task is to help make predictions on medical data. Machine learning is an emerging scientific field in data science dealing with the ways in which machines learn from experience. The aim of this project is to develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques. This project aims to predict diabetes via multiple different supervised machine learning methods including: SVM, Logistic regression, ANN. This project also aims to propose an effective technique for earlier detection of the diabetes disease.*

*Key Words*: Diabetes Prediction, Machine Learning, Deep Learning, Statistical Analysis, Data Visualization, Prediction Model

## 1. INTRODUCTION

People's daily diets contain large amounts of sugar and fat. Due to these factors diabetes risk have increased among people worldwide. As a result, many individuals visit health centers to take related blood tests. However, many of them may not even have the minimum probability of having diabetes. These tests take a lot of time and waste budget of health care systems and people every year. There are researches in progress to provide new ways to help diagnosis of diabetes faster and cheaper to prevent people with least risk of diabetes to get blood tests.

Machine learning techniques are widely used in medical predictions. The learning algorithms use recorded datasets of former patient's information to prepare a model and then use this model with information of an unseen patient to predict if the patient has the desired disease or not.

The aim of this study is to design a more accurate classifier for diabetes predictions. For this reason, we are using machine & deep learning algorithm that improve performance of prediction system in terms of time, cost and accuracy.

## 2. LITERATURE SURVEY

[1] Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers

Diabetes, also known as chronic illness, is a group of metabolic diseases due to a high level of sugar in the blood over a long period. The risk factor and severity of diabetes can be reduced significantly if the precise early prediction is possible. In this literature, we are proposing a robust framework for diabetes prediction where the outlier rejection, filling the missing values, data standardization, feature selection, K-fold cross-validation, and different Machine Learning (ML) classifiers (k-nearest Neighbour, Decision Trees, Random Forest, AdaBoost, Naive Bayes, and XGBoost) and Multilayer Perceptron (MLP) were employed.

[2] Classification of Diabetes using Deep Learning

Deep Learning (DL) is a research area that has different kinds of activation functions and their efficiency is flourished significantly in recent years and has shown remarkable reported by comparative analysis. Potential for artificial intelligence in the field of medical applications. The rest of paper is illustrated in respective manner: We have implemented the DL algorithm for the diabetes basics and background of deep learning techniques is declassification. This paper applied the Multi-Layer Feed Forward Neural Networks (MLFNN) for the diabetes classification.

[3] Deep Belief Neural Network Model for Prediction of Diabetes Mellitus

Diabetes Mellitus is metabolic chronic disease in which blood glucose levels are too high. In India nearly 8.7% of population suffers from diabetes in age range from 20 to 70. Unidentified and untreated diabetes leads to so many health difficulties such as damage of heart, kidneys, eyes, nerves and blood vessels. There are already several methods exists to support clinical decision making but still need improvements to solve the issues and challenges.

[4] A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques

With the continuing increase in the number of the deadly diseases that threaten both human health and life, medical Decision Support Systems (DSS) continue to prove their effectiveness in providing physicians and other healthcare professionals with support in clinical decision making. Among these dangerous diseases, diabetes continues to be one of the leading one that has caused several deaths in the world. It is characterized by an increase in blood sugar levels which can have severe effects on other human organs.

## 3. PROPOSED SYSTEM

Classification is one of the most important decision making techniques in many real world problem. In this work, the main objective is to classify the data as diabetic or non-diabetic and improve the classification accuracy. For many classification problem, the higher number of samples chosen but it doesn't leads to higher classification accuracy. In many cases, the performance of algorithm is high in the context of speed but the accuracy of data classification is low. The main objective of our model is to achieve high accuracy. Classification accuracy can be increase if we use much of the data set for training and few data sets for testing. This survey has analyzed various classification techniques for classification of diabetic and non-diabetic data. Thus, it is observed that techniques like Support Vector Machine, Logistic Regression, and Artificial Neural Network are most suitable for implementing the Diabetes prediction system.
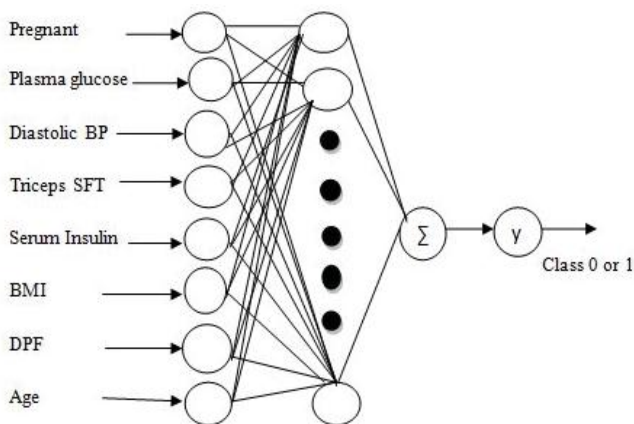


**Fig.1:** System Architecture

## 4. CONCLUSION

Diabetes is the most common disease in the world. In our study, we compare several algorithms with different pre-processing techniques and identify algorithms best performance in which pre-processing technique. We found many of machine learning will give us best accuracy than any other methods. In future we will apply more advanced tricks in Neural Network, such as more hidden layers, algorithm optimization would be more accurate in this case.

## REFERENCES

[1] Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers

[2] Classification of Diabetes using Deep Learning

[3] Deep Belief Neural Network Model for Prediction of Diabetes Mellitus

[4] A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques

[5] A. Misra, H. Gopalan, R. Jayawardena, A. P. Hills, M. Soares, A. A. RezaAlbarrán, and K. L Ramaiya, "Diabetes in developing countries," Journal of Diabetes, vol. 11, no. 7, pp. 522-539, Mar. 2019.

[6] P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karurangaa, N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala, K. Ogurtsova, J. E. Shaw, D. Bright, R. Williams, and IDF Diabetes Atlas Committee, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation," Diabetes Research and Clinical Practice, vol. 157, pp. 107843, Nov. 2019.

[7] M. Maniruzzaman, M. J. Rahman, M. A. M. Hasan, H. S. Suri, M. M. Abedin, A. El-Baz, and J. S. Suri, "Accurate diabetes risk stratification using machine learning: role of missing value and outliers," Journal of Medical Systems, vol. 42, no. 5, pp. 92, May 2018.

[8] N. H. Choac, J. E. Shaw, S. Karuranga, Y. Huang, J. D. R. Fernandes, A. W. Ohlrogge, and B. Malandaa, "IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045," Diabetes Research and Clinical Practice, vol. 138, pp. 271-281, Apr. 2018.

[9] R. Vaishali, R. Sasikala, S. Ramasubbareddy, S. Remya, and S. Nalluri, "Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset," in Proc. International Conference on Computing Networking and Informatics, Oct. 2017, pp. 1-5.