

Research Article



INTERNATIONAL RESEARCH JOURNAL OF PHARMACY

www.irjponline.com

ISSN 2230-8407 [LINKING]

PREDICTION AND ANALYSIS OF DIABETES USING LEARNING ALGORITHMS ON THE DIABETES DATA AND TO COMPARATIVELY ANALYSE THE ALGORITHMS USING INTELLIGIBLE MACHINE LEARNING IN A POPULATION

Dr. Manisha Panda,¹ Dr. Venkat Lakshmi Gondu^{2*}

¹Assistant Professor, Department of Community Medicine, JIS School of Medical Science & Research, Kolkata, West Bengal

^{2*}Assistant Professor, Department of Community Medicine, Shri Rawatpura Sarkar Institute of Medical Sciences and Research, Raipur, Chhattisgarh

Corresponding Address

Email: dr.lakshmigv@gmail.com

How to cite: Panda M, Gondu VL. Prediction and analysis of diabetes using learning algorithms on the diabetes data and to comparatively analyse the algorithms using intelligible machine learning in a population. International Research Journal of Pharmacy, 2024,15:2:8-11.

Doi: 10.56802/2230-8407.1304202

ABSTRACT

Background: Machine learning has become more and more necessary in the last several years to achieve decision-making patterns and intriguing judgments in practically every industry. This has also been used to analyse health data, where several methods are used to evaluate the data. These days, health data is precise, sensitive, and essential, necessitating precise analysis with outcomes that are critical and paramount. Machine learning has improved data analysis, sensitivity, role, and interest.

Aim: The purpose of this study was to use learning algorithms to analyse and forecast diabetes utilizing diabetic data, as well as to compare the algorithms.

Methods: The median approach was employed for preprocessing the dataset in this study. After preprocessing, the diabetes dataset in the current study was subjected to ten distinct machine learning methods.

Results: In order to predict diabetes, the current study employed a diabetes dataset with eight characteristics or symptoms. In order to improve the classification method, several machine learning outcomes of the processes were examined and contrasted. The data from this study can be used in future research projects.

Conclusion: The current study comes to the conclusion that, in comparison to other machine learning procedures, linear support vector machines produce superior detection outcomes.

Keywords: Diabetes Prediction, Decision Tree, Gaussian Process, Machine Learning, Nearest Neighbor, Predictive Analysis, SVM,

INTRODUCTION

The health care industry is concerned about the rising incidence of diabetes worldwide. Diabetes is challenging to predict, test for, and detect. In order to determine the pattern required universally, machine learning has lately become widely applied in the medical profession. According to estimates, 422 million people worldwide had diabetes in 2014, and the prevalence of the disease among adults increased from 4.7% in 1980 to 8.5% in 2014.¹

Diabetes was directly responsible for an estimated 1:6 million deaths globally in 2015 alone. Furthermore, compared to a non-diabetic individual, a diabetic patient has an increased risk of cardiovascular disease, vision impairment, and limb amputation. Diabetes early diagnosis, intervention, and prevention have become a major worldwide health problem because to the significant socioeconomic costs that the disease places on affected families as well as the local healthcare system.²

The World Health Organization and others consider impaired glucose tolerance (IGT), which defines the body's aberrant insulin response, to be one of the most significant risk factors. WHO and the American Diabetes Association (ADA), for identifying pre-diabetes, the kind of diabetes that is detected early.³

The glucose clamp technique is a dangerous experiment that needs highly skilled staff, which restricts its application in clinical practice or big epidemiological investigations. Nevertheless, it can quantify the IGT. An oral glucose tolerance test (OGTT) is a less intrusive method of measuring the IGT that measures blood glucose and insulin levels in response to a standardized glucose dosage administered orally two hours before to the testing, following an overnight fasting condition.⁴

The health sector's sensitivity forces academics to concentrate on machine learning. Many fields employ machine learning to address data-driven problems. One Machine learning implementation necessitates data expertise and yields varying outcomes across datasets. A dataset may contain null values, inaccurate values, and incomplete data. Additionally, data preparation is necessary to get good results since it enhances performance and aids in better prediction of results. When evaluating data, the median approach is typically employed.⁵

Predicting diabetes is crucial since it is an incurable condition. With the right diet, exercise, and care, side effects and problems can be minimized and controlled. The illness has detrimental repercussions and might possibly result in death. Various techniques may be employed with machine learning, and classifiers' potential can be compared based on how accurate they are.⁶ Algorithms for machine learning classification may also be applied. Deep learning techniques have been applied in the past to accurately and effectively analyse health care data using computation methods of both distributed and parallel systems.⁶

Other research has examined machine learning and trunk sway data analysis methods to provide a fair, automated assessment and offer accurate analysis outside of clinical settings. The health care system can make use of systemic techniques and algorithms that are independent of one another. Earlier, a machine learning-based strategy for lowering postprandial glucose levels was also put out. This machine learning method has shown promise in a number of areas.⁷

Investigating the application of advanced machine learning techniques to create customized models with the goal of identifying risk factors for Type 2 diabetes mellitus for the incidence of both fatal and non-fatal cardiovascular disease was the goal of the current study. Additionally, the study employs deep learning to build the association between bad effects and laboratory and medical parameters, with an emphasis on public health and medical imaging.

MATERIALS AND METHODS

Investigating the application of advanced machine learning techniques to create customized models with the goal of identifying risk factors for Type 2 diabetes mellitus for the incidence of both fatal and non-fatal cardiovascular disease was the goal of the current study.

The project also employs deep learning to explore the association between undesirable effects and laboratory and medical parameters, with an emphasis on public health and medical imaging. Researchers may concentrate on studying the data to identify and forecast the illness because of the increased prevalence of diabetes and the abundance of available data. Utilizing diabetes data, the current study aids in illness prediction by utilizing disease characteristics.

A vast number of people are affected by the rising prevalence of diabetes. When not properly managed, it has a detrimental impact on life and can possibly result in death. Because diabetes cannot be cured, it is crucial to forecast its course. The study interprets the findings using diabetes prediction algorithms. The institution dataset was used in the research. The research was carried out, at Department of Community Medicine, JIS School of Medical Science & Research, Kolkata, West Bengal after obtaining clearance from the concerned ethical committee. The study makes predictions about a diabetic dataset based on its properties.

The current dataset makes use of several records and various symptoms, including pregnancy, blood pressure, glucose, BMI, insulin, skin thickness, age, pedigree function, and skin thickness. The signs of an illness are used to forecast its course. Noisy data, irrelevant data, and null values can all lead to inaccurate predictions and poor decisions. To prevent these issues, preparation of the data is carried out.

Preprocessing is done to achieve the right values, which improves the outcomes compared to the raw data. Cleaning, integrating, transforming, and reducing data are all part of pre-processing. Learning models are utilized for diabetes prediction. Prior to knowing how to use an algorithm, preprocessing is required.

All eight parameters in the current study include null values, and missing values were substituted with median values. Following that, no null values were left for any of the following parameters: age, pedigree function, blood pressure, skin thickness, insulin, BMI, glucose, pregnancies, and skin thickness.

Eight characteristics are used in the diabetes dataset, and a variety of machine learning models, such as QDA, Naïve Bayes, Adaboost, MLP classifier, Random Forest, Decision tree, Gaussian process, RBF SVM, Linear SVM, and Nearest Neighbor, are used for analysis.

RESULTS

The goal of this study was to find out how advanced machine learning techniques may be used to create customized models that would identify risk variables for Type 2 diabetes mellitus in relation to the occurrence of both fatal and non-fatal cardiovascular disease.

With an emphasis on medical imaging and public health, the study also employs deep learning to build the association between laboratory and medical parameters and bad consequences. To identify diabetes and take the necessary measures in a timely manner, the strength of eight symptoms was employed. Because diabetes cannot be cured, it requires special care. It is essential for diabetes detection.

The diabetes dataset was pre-processed using a variety of learning techniques, and then it was exposed to ten distinct learning paradigms. The accuracy of these models was then used to evaluate their potential. The current study's findings demonstrated that various machine learning models produced varying levels of accuracy.

According to Table 1, the accuracy of the (Quadratic Discriminant Analysis) QDA classifier was 73.93, the Naive Bayes classifier was 73.77, the Adaboost classifier was 73.1, the (Multi-Layer) MLP classifier was 67.4, the Random Forest classifier was 74.7, the Decision Tree was 70.01, the Gaussian Process was 67.05, the RBF (Radial basis function) Support Vector Machine was 65.29, the Linear Support Vector Machine was 77.01, and the Nearest Neighbor was 71.15.

DISCUSSION

The current study looked at the use of advanced machine learning techniques to create customized models with the goal of identifying risk variables for Type 2 diabetes mellitus in relation to the occurrence of both fatal and non-fatal cardiovascular disease.⁶ With an emphasis on medical imaging and public health, the study also employs deep learning to build the association between laboratory and medical parameters and bad consequences. The strength of 8 symptoms was used for diabetes detection to take needed precautions timely. Because diabetes cannot be cured, it requires special care. It is essential for diabetes detection.⁷

The diabetes dataset was pre-processed using a variety of learning techniques, and then it was exposed to ten distinct learning paradigms. The accuracy of these models was then used to evaluate their potential. The current study's findings demonstrated that various machine learning models produced varying levels of accuracy.⁸ The accuracy was 73.93 with the QDA classifier, 73.77 with the Naive Bayes, 73.1 with Adaboost, 67.4 with the Multi-Layer MLP classifier, 74.7 with Random Forest, 70.01 with the Decision Tree, 67.05 with the Gaussian Process, 65.29 with the RBF (Radial basis function) Support Vector Machine, 77.01 with the Linear Support Vector Machine, and 71.15.⁸ with the nearest neighbour.

When it comes to detection, the linear support vector machine performs better than the others. Since linear SVM⁹ had more accuracy than other approaches, it showed greater promise with the diabetes dataset. Additionally, the classifier performed well on the current dataset.⁹ With eight factors evaluated, the linear model produced a better output with better outcomes.

CONCLUSION

Despite its limitations, the current study comes to the conclusion that diabetes necessitates early detection and prediction since it places a significant strain on the Indian healthcare system. With the most recent machine learning protocols and algorithms, which provide improved detection with linear support vector machines, high illness sensitivity enhances its evaluation.

REFERENCES

- 1) Lazar A. A comparison of linear support vector machine algorithms on large non-sparse datasets. In 2010 Ninth International Conference on Machine Learning and Appl. 2010 Dec 12 (pp. 879-882). IEEE.

- 2) Barakat N, Bradley AP, Barakat MN. Intelligible support vector machines for diagnosis of diabetes mellitus. *Transac Inform Tech Biomed* 2010;14:1114-20.
- 3) Jaya R, Kumar SM. A Study on Data Mining Techniques Methods Tools and Applications in Various Industries. *Int J Curr Res Rev* 2016;8:34.
- 4) Abbas H, Alic L, Rios M, Abdul-Ghani M, Qaraq K. Predicting diabetes in healthy population through machine learning. In 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) 2019:567-70.
- 5) Rajesh R, Singh SA, Vaithy KA, Manimekalai K, Kotasthane D, Rajasekar SS. The effect of mucuna pruriens seed extract on pancreas and liver of diabetic Wistar rats. *Int J Curr Res Rev* 2016;8:61.
- 6) Cao W, Czarnek N, Shan J, Li L. Microaneurysm detection using principal component analysis and machine learning methods. *Transac Nanobiosci* 2018;17:191-8.
- 7) Zarkogianni K, Athanasiou M, Thanopoulou AC, Nikita KS. Comparison of machine learning approaches toward assessing the risk of developing cardiovascular disease as a long-term diabetes complication. *J Biomed Health Infor* 2017;22:1637-47.
- 8) Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ. Deep learning for health informatics. *J. Biomed Health Infor* 2016;21:4-21.
- 9) Wagh SP, Bhagat SP, Bankar N, Jain K. Role of Vitamin-C Supplementation in Type II Diabetes Mellitus. *Int J Curr Res* 2020;12:61-64.

Machine Learning Models	Accuracy
(Quadratic Discriminant Analysis) QDA classifier	73.93
Naive Bayes	73.77
Adaboost	73.1
(Multi Layer) MLP classifier	67.4
Random Forest	74.7
Decision Tree	70.01
Gaussian Process	67.05
RBF (Radial basis function) Support Vector Machine	65.29
Linear Support Vector Machine	77.01
Nearest Neighbor	71.15

Table 1: Accuracy seen with various machine learning models