Research Article

INTERNATIONAL RESEARCH JOURNAL OF PHARMACY

www.irjponline.com

ISSN 2230-8407 [LINKING]

DIAGNOSIS OF DIABETES USING INTELLIGIBLE MACHINE LEARNING IN A POPULATION

Dr. Shiv Kumar,[1*] Dr. Swayam Prakash[2]

[1*]Associate Professor, Department of General medicine, I care Institute of Medical Sciences and research & Dr. Bidhan Chandra Roy Hospital, Haldia, West Bengal

[2]Assistant Professor, Department of General medicine, I care Institute of Medical Sciences and research & Dr. Bidhan Chandra Roy Hospital, Haldia, West Bengal

Address for correspondence

Dr. Shiv Kumar

Email: drkumarshiv1974@gmail.com

How To Cite: Kumar S, Prakash S. Diagnosis of Diabetes using intelligible machine learning in a population. International Reasearch Journal of Pharmacy, 2022, 13:8:24-27.

DOI: 10.32937/2230-8407.3744

==========

ABSTRACT

Introduction: In the recent past, to attain decision-making patterns and interesting decisions, in almost every field, demand for machine learning has increased. This has also been applied to health data where effective analysis uses different techniques to assess the data. Presently, health data is accurate, crucial, and sensitive requiring accurate analysis, where results achieved are vital and of prime importance. Machine learning has increased sensitivity, role, interest, and data analysis.

Objectives: The present study was conducted to predict and analyze diabetes using learning algorithms on the diabetes data and to comparatively analyze the algorithms.

Methods: In the present study, for preprocessing the dataset median method was used. Following the preprocessing, 10 different algorithms from machine learning were used and applied to the dataset of diabetes in the present study.

Results: The present study used a diabetes dataset having 8 features or symptoms to predict diabetes. Different machine learning results of the mechanisms were analyzed and compared, to attain a better classification technique. In researches conducted in the future, the data from the present study can be utilized.

Conclusion: The present study concludes that better detection results are seen with linear support vector machines compared to other machine learning processes.

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

INTRODUCTION

Increasing diabetes prevalence globally is a matter of concern in the health care sector. Prediction, screening, and detection of diabetes is difficult and is concerning. Machine learning is recently used widely in the field of medicine to assess the pattern needed ubiquitously. Sensitivity in the health sector pushes the researchers to focus on machine learning. Machine learning is used in many areas to work with issues that are data-driven.[1] Implementation of machine learning requires data knowledge and shows different results in a different dataset. In a dataset, there may be incomplete data, incorrect values, and null values. Also, data preprocessing is needed to achieve good results as it improves performance helping in better results prediction. The median method is usually used to assess data.[2]

As diabetes is an incurable disease, its prediction is essential. Ill effects and complications can be controlled and reduced with precautions, proper diet, and exercise. The disease can even lead to death with negative effects. With machine learning, different algorithms can be used and the potential of classifiers can be compared depending upon their accuracy.[3] Machine learning classification algorithms can also be used. In the past, computation techniques of

distributed as well as parallel systems have been used techniques of deep learning to properly and efficiently analyze the data related to health care.[4]

Other literature data has analyzed techniques of machine learning and trunk sway data to get a balanced and automatic evaluation and provide correct analysis outside the clinics. Algorithm independent across algorithms and systemic approaches can be used in the health care system. Previously, a method for the reduction of postprandial glucose-regulated with machine learning is also been proposed. This approach of machine learning is proved useful in various applications.[5]

The present study was conducted to investigate the use of sophisticated machine learning techniques to develop personalized models having the target on detection of risk factors of Type 2 diabetes mellitus for fatal and non-fatal cardiovascular disease incidence. The study also focuses on developing the relationship between laboratory and medical factors and adverse effects using deep learning with a focus on medical imaging and public health.

MATERIALS AND METHODS

The present study was conducted to investigate the use of sophisticated machine learning techniques to develop personalized models having the target on detection of risk factors of Type 2 diabetes mellitus for fatal and non-fatal cardiovascular disease incidence. The study also focuses on developing the relationship between laboratory and medical factors and adverse effects using deep learning with a focus on medical imaging and public health. Increased diabetes burden and availability of extensive data help researchers to focus on analyzing it to detect and predict the disease. The present study uses diabetes data and helps in disease prediction based on disease features.

Diabetes prevalence is increasing and affecting a large population. It negatively affects life and even leads to death when not adequately controlled. Diabetes prediction is important as it is incurable. The study uses the prediction algorithms on diabetes for results interpretation. The study utilized the institution dataset. The study was conducted at…..from….to….after obtaining clearance from the concerned ethical committee. The study uses a diabetes dataset and based on features this dataset prediction is done.

The present dataset utilizes many records and different symptoms, age, pedigree function, BMI, insulin, skin thickness, blood pressure, glucose, and pregnancy. Disease prediction is done based on the disease symptoms. Incorrect prediction and wrong judgment can result from the data can result from noisy data, non-relevant data, and null values. Hence, preprocessing of the data is done to avoid these complications.

Preprocessing is done to attain correct values helping in getting results better than the data which was not processed. Preprocessing includes data cleaning, integration, transformation, and reduction. For diabetes prediction, learning models are used. Before learning algorithm application, preprocessing is necessary. Null values are seen in the present study for all 8 parameters, also included missing values which were replaced by the median values. No null values were then left for any parameter including pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, pedigree function, and age.

Diabetes dataset uses the data of 8 features and applies various machine learning models for analysis including QDA, Naïve Bayes, Adaboost, MLP classifier, Random Forest, Decision tree, Gaussian process, RBF SVM, Linear SVM, and Nearest Neighbor.

RESULTS

The present study was conducted to investigate the use of sophisticated machine learning techniques to develop personalized models having the target on detection of risk factors of Type 2 diabetes mellitus for fatal and non-fatal cardiovascular disease incidence. The study also focuses on developing the relationship between laboratory and medical factors and adverse effects using deep learning with a focus on medical imaging and public health. The strength of 8 symptoms was used for diabetes detection to take needed precautions timely. Special attention is needed for diabetes owing to its incurable nature. It is vital for detecting diabetes.

Various learning algorithms were used in the diabetes dataset following preprocessing, and the dataset was then subjected to 10 different learning paradigms where the potential of these models was assessed based on accuracy. The results of the present study have shown that there was different accuracy using different machine learning models. The accuracy with (Quadratic Discriminant Analysis) QDA classifier was 73.93, with Naive Bayes was 73.77, Adaboost was 73.1, (Multi-Layer) MLP classifier was 67.4, Random Forest was 74.7, Decision Tree was 70.01, Gaussian Process

was 67.05, RBF (Radial basis function) Support Vector Machine was 65.29, with Linear Support Vector Machine was 77.01, and with Nearest Neighbor 71.15 as shown in Table 1.

DISCUSSION

The present study was conducted to investigate the use of sophisticated machine learning techniques to develop personalized models having the target on detection of risk factors of Type 2 diabetes mellitus for fatal and non-fatal cardiovascular disease incidence.[6] The study also focuses on developing the relationship between laboratory and medical factors and adverse effects using deep learning with a focus on medical imaging and public health. The strength of 8 symptoms was used for diabetes detection to take needed precautions timely. Special attention is needed for diabetes owing to its incurable nature. It is vital for detecting diabetes.[7]

Various learning algorithms were used in the diabetes dataset following preprocessing, and the dataset was then subjected to 10 different learning paradigms where the potential of these models was assessed based on accuracy. The results of the present study have shown that there was different accuracy using different machine learning models.[8] The accuracy with (Quadratic Discriminant Analysis) QDA classifier was 73.93, with Naive Bayes was 73.77, Adaboost was 73.1, (Multi-Layer) MLP classifier was 67.4, Random Forest was 74.7, Decision Tree was 70.01, Gaussian Process was 67.05, RBF (Radial basis function) Support Vector Machine was 65.29, with Linear Support Vector Machine was 77.01, and with Nearest Neighbor 71.15.[8]

Linear support vector machine shows a better detection result compared to others. As the accuracy was better with linear SVM19 compared to other methods, it showed better potential with the diabetes dataset. Also, the classifier was good for the present dataset.[9] The linear model provides a better output which showed better results with 8 parameters assessed.

CONCLUSION

Within its limitations, the present study concludes that diabetes poses a high burden on the Indian healthcare system hence requires early identification and prediction. High disease sensitivity increases its assessment with the latest machine learning protocols and algorithms showing better detection with linear support vector machines.

REFERENCES

1) Lazar A. A comparison of linear support vector machine algorithms on large non-sparse datasets. In2010 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, Qaraqe K. Predicting diabetes in healthy population through machine learning. In2019 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.

TABLES

S. No	Machine Learning Models	Accuracy
1)	(Quadratic Discriminant Analysis) QDA classifier	73.93
2)	Naive Bayes	73.77
3)	Adaboost	73.1
4)	(Multi Layer) MLP classifier	67.4
5)	Random Forest	74.7
6)	Decision Tree	70.01
7)	Gaussian Process	67.05
8)	RBF (Radial basis function) Support Vector Machine	65.29
9)	Linear Support Vector Machine	77.01
10	Nearest Neighbor	71.15

Table 1: Accuracy seen with various machine learning models