

Analysis of Various Diabetes Prediction Techniques Based on Machine Learning

GurmanvirKaur^{1,*}, Tejbir Singh²

Abstract

Data mining is a method of mine from the underdone data for maintains useful information. In organize to obtain fundamental knowledge it is important to remove bulky quantity of data. This method is consist of many important components like data cleaning, data selection, data integrity, pattern evaluation with data mining engine and database with graphical user interface. Diabetes is a chronic disease that is also known as Non-Insulin Dependent Diabetes Mellitus, or Adult Onset Diabetes Mellitus. The adequate insulin is created by the patient, which can't be used by body because of absence of affectability to insulin by the cells of the body. The various techniques for the diabetes prediction are reviewed in terms of certain parameters. Many methods of machine learning is used on data mining for primary knowledge like supervised, unsupervised, deep and reinforcement learning techniques. KNN is a part of supervised learning and ANN is an example of deep learning. Neural network is a part of machine learning that can be used for measurement to every disease on human body. If any disease present on human body that indicate in machine. This paper is very important to human body area network. This network is used for analysis to every disease in human body.

Keywords: Machine Learning, Diabetes prediction, Feature extraction, KNN, ANN, KDD

INTRODUCTION

Diabetes is a frequent chronic malady. This disease severely affects the health of a human being. The increase in blood sugar level from the normal range is the main feature of this disease. Imperfect insulin secretion or impaired genetic effects are the main causes of this disease. The first type generally affects youngsters below thirty year. Some medical indications of this disease are more thirst and repeated urination, high blood sugar levels etc. it is not possible to cure this disease with oral medicines only. In many cases, insulin is given to the human body through injection. The second type of this disease mainly occurs in middle-aged and old people. In old people, this disease mainly occurs due to fatness, high blood pressure, dyslipidemia, arteriosclerosis, and other maladies. This disease has become very common in people's day to day life with the improved standard of living [1]. Hence, quick and accurate diagnosis and analysis of this disease is a major field of research. Some parameters such as blood sugar in fasting, sugar tolerance, and arbitrary blood sugar levels are used in

medical field for the diagnosis of this disease. It is possible to control this disease by diagnosing it in early stage. On the other hand, it is a complex job for medical practitioners to predict this disease in early stage. This is because of the compound interdependence of this disease on several aspects.

Data Mining can be defined as a technology using which valuable information can be extracted from the massive volume of data. The big patterns can be explored and analyzed using statistical and Artificial Intelligence in big databases. The future trends can

*Author for Correspondence

Gurmanvir Kaur,
E-mail: gurmanvirkaur4@gmail.com

¹M.Tech Student, Department of Computer Science and Engineering, GNI Mullana, Haryana, India

²Assistant Professor, Department of Computer Science and Engineering, GNI Mullana, Haryana, India

Received Date: March 17, 2021

Accepted Date: March 22, 2021

Published Date: April 15, 2021

Citation: GurmanvirKaur, TejbirSingh. Analysis of Various Diabetes Prediction Techniques Based on Machine Learning. Journal of Artificial Intelligence Research & Advances. 2021; 8(1): 1–5p.

be predicted or hidden pattern can be discovered using data mining. Data-mining methods have been extensively utilized for predicting blood sugar levels. The data-mining methods do not need strong model suppositions for making prediction models for blood sugar levels. Data mining has the ability to get subtle underlying patterns and associations in experiential data [2]. Therefore, data mining provides efficiently predicts the sugar level within blood. In general, different studies have used data-mining methods for predicting blood sugar levels with and without fasting. There are several steps included in data mining for diabetes disease prediction. These steps are Data pre- processing, Data reduction, Classification, Performance analysis. Data pre-processing is the first step in diabetes prediction. In this step, the rough data is converted into a reasonable format. The data obtained from real-world is often imperfect, incompatible, and/or short of certain behaviors or trends. This data has many errors. Data pre-processing approach efficiently resolves these issues. This approach makes rough data ready for more processing. Data quality is the main aspect of the entire prediction model. This parameter may affect the prediction result provided by analysis. Therefore, it is necessary to carry out data pre-processing prior to analyze data [3]. In data reduction task, the implementation of a clustering algorithm is done prior to the use of classification approach for removing the improperly classified patterns. Clustering discovers the wrongly classified instances. Classification is a data mining technique. This technique allocates objects in a compilation for targeting classes. The main aim of this technique is to make prediction about the target class for every case in the data. As an instance, it is possible to use a classification model for classifying loan applicants a slow, medium, or high credit risks. This task initiates with a data set with specified class allocations. A classification model for predicting credit risk could be designed on the basis of experiential data for various loan applicants over duration [4]. The testing of classification models is carried out by comparing the forecasted values to recognized target values in a test dataset. In feature evaluation task, several measures are applied to evaluate how efficiently or accurately a classification model predicts the class label of tuples. There are various data mining algorithms which are used for diabetes prediction. Some popular data mining algorithms for diabetes prediction include Decision Tree, Random Forest, Neural Network, and SVM (Support Vector Machine). Decision tree is a fundamental classification and regression algorithm [5]. Decision tree model has a tree like configuration. This configuration explains the procedure of classification instances on the basis of attributes. This algorithm can be regarded as a set of if-then rules. This algorithm may also be considered as conditional probability allocations described in feature location and class location. The tree in these classifiers starts with a sole node. This node represents the training patterns. The node becomes the leaf node if the patterns occur in the similar class. The class marks this node as leaf node. In other case, the algorithm makes the selection of inequitable feature as the existing node of the decision tree. The division of training samples is done into a number of subsets as per the value of the existing decision node feature. Every subset makes a branch. There are different rules for different branches. The repletion of earlier steps is carried out in recursive manner for every subset or branch formed in the earlier step. This phenomenon generates a decision tree on every divided pattern. Random Forest classifier uses various trees to carry out classification. This algorithm is developed by Breiman. It is a multifunctional machine learning technique. This algorithm can carry out both prediction as well as regression. Moreover, this classification algorithm depends on Bagging. An important role has been played by this algorithm in ensemble machine learning. Various research works based on biomedicine use this approach. This classifier produces various decision trees [6]. Hence, this classifier is not similar to decision tree algorithm. Every tree in this classifier generates its own classification output and ‘vote,’ when this classifier [7].

LITERATURE REVIEW

Desmond BalaBisandu, et.al (2020) carried out a research work which was focused on the analysis predictive model for the prediction of T2DM (Type 2 diabetes mellitus) [8]. The work was concerned with developing an analytical predictive model capable to guide decision making process of medical personnel more successfully. This work also provided understanding about the diabetes disease. In this work, the Artificial Neural Network (ANN) model had been analyzed and then enhanced by

inserting more affecting aspects and evaluating the features. It was concluded that the devised machine learning model for predicting and detecting diabetes mellitus is imperative due to the continuous increase in the seriousness of condition globally with the increased financial cost.

K. Lakshmi Priya, et.al (2020) conducted a research work with the aim to analyse the diabetic patients' datasets along with predicting the diabetes mellitus disease in the initial stage [9]. This work made use of Naïve Bayes classifier model for diabetes prediction. The information from medical databases, carrying 1865 incidences with different features was gathered to arrange information in the diabetes dataset of patients. The tested outcomes revealed that the applied new approach had the potential to predict diabetes with better accuracy rate of 96% in comparison to the state-of-the-art methods. The recommended scheme based on Naïve Bayes Classifier generated predictive results of diabetes by accepting different input features such as blood pressure, sugar level, etc. Predictions based on all these factors were highly accurate.

Zhongxian Xu, et.al (2019) suggested the use of a risk prediction model for type 2 diabetes prediction [10]. The recommended model was based on Weighted Feature Selection of Random Forest (RF-WFS) and XGBoost (extreme gradient boosting) Ensemble Classification model. The recommended model used RF-WFS and XGBoost (extreme gradient boosting) classifiers for selecting best features. A lot of performance parameters were compared in this work to validate the efficiency of the recommended classifiers. The achieved results demonstrated that the recommended model outperformed the other existing prediction models in terms of classification accuracy. These results established the efficacy of the recommended model in the early detection of diabetes disease.

B.V. Baiju, et.al (2019) provided a comprehensive review of data mining-based disease prediction algorithms using the database of diabetes disease [11]. The variation in the prediction results was noticed with different types of employed techniques, events and attributes. Diabetic prediction based on a DIM (Disease Influence Measure) was recommended in this work for predicting diabetes. At first, the recommended technique carried out the preprocessing of input data set and eliminated the noisy features. This technique estimated disease influence measure (DIM) on the basis of input data point attributes in the next step. The recommended technique predicted diabetes disease on the basis of DIM value. Various algorithms for predicting diabetes disease were considered in this work. These algorithms were compared on the basis of their performance in diabetes forecasting.

Fikirte Girma Woldemichael, et.al (2018) suggested the use of data mining algorithms for diabetes prediction. Back propagation algorithm was recommended in this work to classify people as diabetic or normal [12]. Different machine algorithms were utilized in this work for predicting diabetes disease. These neural networks had an input layer with eight parameters. Every hidden layer had six neurons. This network generated one output layer. The performance of the model was improved using fivefold cross-validation method and big value learning rate. PIMA Indian dataset was used in this work. The implementation of this work was done in RStudio with the help of R programming language. The tested results depicted that the recommended approach achieved accuracy, sensitivity and specificity of 83.11 %, 86.53% and 76% respectively. The recommended algorithm outperformed the other existing techniques in diabetes prediction.

Han Wu, et.al (2018) suggested the use of a new prediction model type 2 diabetes mellitus (T2DM) [13]. Various data mining algorithms were used in this work for diabetes prediction. The main aim here was to enhance the accuracy of the prediction model and making it suitable for several datasets. This model was divided into two segments on the basis of a sequence of preprocessing events. These segments were identified as improved K-means algorithm and the logistic regression algorithm. Two data sets were used in this work for comparing the results of this work and the results of other researches. The achieved results depicted that the recommended model showed 3.04% more prediction accuracy than other existing models. In addition, the recommended model made certain the

lack of data quality. The implementation of recommended model was done two other diabetes datasets for further evaluating the performance of the recommended model. The recommended model performed well in both cases. Hence, this model was very efficient in the real time handling of diabetes disease.

DeerajShetty, et.al (2017) stated that the main aim of data mining was to extract data from the database and convert this data into valuable information [14]. This research work was mainly focused on the use of data mining technology in the healthcare domain. The use of this technology in medical field proved advantageous for both physicians and patients. An Intelligent Diabetes Disease Prediction model was developed in this Comparison Table 1 work for diabetes prediction. This model analyzed the diabetes disease using the dataset of diabetic patients. Two data mining algorithms called Bayesian and KNN (K-Nearest Neighbor) were implemented on the dataset of diabetic patients. Several features of diabetes were extracted from the dataset for analyzing the performance of these two algorithms in diabetes prediction.

Table 1. Literature Review.

Author	Year	Description	Outcome
LinditaLoku, BekimFetaji, MajlindaFetaji,	2020	Caried out a research work which was focused on the analysis predictive model for the prediction of T2DM (Type 2 diabetes mellitus). The work was concerned with developing an analytical predictive model capable to guide decision making process of medical personnelmore successfully.	It was concluded that the devised machine learning model for predicting and detecting diabetes mellitus is imperative due to the continuous increase in the seriousness of condition globally with the increased financial cost.
K. Lakshmi Priya, MouryaSaiCharan Reddy Kypa, MuchumarriMadhuSudhan Reddy, G. Ram Mohan Reddy	2020	Conducted a research work with the aim to analyse the diabetic patients' datasets along with predicting the diabetes mellitus disease in the initial stage. This work made use of Naïve Bayes classifier model for diabetesprediction.	The recommended scheme based on Naïve Bayes Classifier generated predictive results of diabetes by accepting different input features such as blood pressure, sugar level, etc. Predictions based on all thesefactors were highly accurate.
ZhongxianXu,ZhiliangWang	2019	Suggested the use of a risk prediction model for type 2 diabetes prediction. The recommended model was based on Weighted Feature Selection of Random Forest (RF-WFS) and XGBoost (extreme gradient boosting)EnsembleClassification model.	The achieved results demonstrated that the recommended model outperformed the other existing prediction models in terms of classification accuracy. These results established the efficacy of the recommended model in the early detection of diabetes disease.
B.V.Baiju,D.John Aravindhar	2019	Provided a comprehensive review of data mining-based disease prediction algorithms using the database of diabetes disease. There were various signs that indicated the occurrenceofthis severe disease.	Various algorithms for predicting diabetes disease were considered in this work. These algorithms were compared on the basis of their performance in diabetes forecasting.
FikirteGirmaWoldemichael, SumitraMenaria,	2018	Suggested the use of data mining algorithms for diabetes prediction. Back propagation algorithm was recommended inthis work to classify people asdiabetic or normal.	The tested results depicted that the recommended approach achieved accuracy, sensitivity and specificity of 83.11 %,86.53%and 76% respectively.
Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, Xiaoyi Wang	2018	The tested results depicted that the recommended approach achieved accuracy,sensitivity and specificity of 83.11 %, 86.53% and 76% respectively.	Theimplementationof recommended model was done two other diabetes datasets for further evaluating the performance of the recommended model. The

			recommended model performed well in both cases. Hence, this model was very efficient in the real time handling of diabetes disease.
Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil	2017	An Intelligent Diabetes Disease Prediction model was developed in this work for diabetes prediction. This model analyzed the diabetes disease using the dataset of diabetic patients.	Several features of diabetes were extracted from the dataset for analyzing the performance of these two algorithms in diabetes prediction.

CONCLUSION

The diabetes prediction is the problem of prediction analysis. The diabetes prediction analysis can be done in two steps which are features extraction and classification. In this work, it is concluded that in the existing work, the SVM classification approach is applied for the prediction analysis. The SVM classifier has less accuracy, as it is not able to drive relationship between the attributes and target set. It is analyzed that novel classification method will be designed for the diabetes prediction.

REFERENCES

1. Dr.D.I.George Amalarethinam, N.Aswin Vignesh. Prediction of Diabetes mellitus using Data Mining Techniques: A Survey. International Journal of Applied Engineering Research.2015.
2. Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, Ioanna Chouvarda. Machine Learning and Data Mining Methods in Diabetes Research. Computational and Structural Biotechnology Journal.2017;15:104–116.
3. Francesco Mercaldo, Vittoria Nardone, Antonella Santone. Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. International Conference on Knowledge based and International Information and Engineering System.2017; 112:2519-2528.
4. Roxana Mirshahvalad, Nastaran Asadi Zanjani. 9th International Conference on Computational Intelligence and Communication Networks: Diabetes Prediction Using Ensemble Perceptron Algorithm; 2017 Sept 16-17; Girne, Northern Cyprus, New York: IEEE; 2018 March 19.
5. Miss. Sneha Joshi, Prof. Megha Borse. International Conference on Micro-Electronics and Telecommunication Engineering: Detection and Prediction of Diabetes Mellitus Using Back-Propagation Neural Network; 2016 22-23 Sept; Ghaziabad, India, New York: IEEE; 2017 June 08.
6. Deepti Sisodia, Dilip Singh Sisodia. Prediction of Diabetes using Classification Algorithms. International Conference on Computational Intelligence and Data Science. 2018; 132:1578-1585.
7. Nongyao Nai-aruna, Rungruttikarn Mounmaia. Comparison of Classifiers for the Risk of Diabetes Prediction. 7th International Conference on Advances in Information Technology.2015; 69: 132-142.
8. Lindita Loku, Bekim Fetaji, Majlinda Fetaji. International Congress on Human-Computer Interaction, Optimization and Robotic Applications: Prevention of Diabetes by Devising A Prediction Analytics Model; 2020 June 26-28; Ankara, Turkey, New York: IEEE; 2020 July 30.
9. K. Lakshmi Priya, Mourya Sai Charan Reddy Kypa, Muchumarri Madhu Sudhan Reddy, G. Ram Mohan Reddy. 4th International Conference on Trends in Electronics and Informatics: A Novel Approach to Predict Diabetes by Using Naive Bayes Classifier. 2020 June 15-17; Tirunelveli, India, New York: IEEE; 2020 July 17.
10. Zhongxian Xu, Zhiliang Wang. Eleventh International Conference on Advanced Computational Intelligence: A Risk Prediction Model for Type 2 Diabetes Based on Weighted Feature Selection of Random Forest and XGBoost Ensemble Classifier; 2019 June 7-9; Guilin, China, New York: IEEE; 2019 July 29.
11. B.V. Baiju, D. John Aravindhar. 1st International Conference on Innovations in Information and Communication Technology: Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining; 2019 April 25-26; Chennai, India, New York: IEEE; 2019 June 20.
12. Fikirte Girma Woldemichael, Sumitra Menaria. Prediction of Diabetes Using Data Mining Techniques. 2nd International Conference on Trends in Electronics and Informatics. 2018.

13. Han Wu, Shengqi Yang, Zhangqin Huang, et al. Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*. 2018; 10: 100-107.
14. DeerajShetty, KishorRit, SohailShaikh et al. International Conference on Innovations in Information, Embedded and Communication Systems: Diabetes disease prediction using data mining; 2017 March 17-18; Coimbatore, India, New York: IEEE; 2018.