

Analysis of Various Diabetic Prediction Techniques

Tejinder Sharma^{1*}, Nitika Sharma²

^{1,2}Department of Computer Science & Engineering, Amritsar Group of Colleges, Punjab, India

*Corresponding Author: tejinder.acet@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v8i12.7073> | Available online at: www.ijcseonline.org

Received: 20/Oct/2020, Accepted: 08/Dec/2020, Published: 31/Dec/2020

Abstract- The data mining is the approach which can mine required information from the rough data. The prediction analysis is the approach which can predict future possibilities based on the current information. This review paper, is based on the diabetic prediction. The diabetic prediction technique has various steps like data pre-processing, feature extraction and classification. In this paper, various diabetic prediction techniques are reviewed and analyzed in terms of certain parameters.

Keywords- Diabetic prediction, classification, feature extraction

I. INTRODUCTION

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 be predicted or hidden pattern can be discovered using data mining. There are different types of data mining techniques. These techniques include classification, clustering, association rule, regression, outlier detection etc. The technology of data mining is gaining a lot of popularity in healthcare sector. Many researchers are implementing data mining techniques in the field of bioinformatics. Bioinformatics can be defined as a science of storing, fetching, arranging, interpreting and using information obtained from biological series and molecules. In the current scenario, Knowledge discovery and data mining techniques are extensively used for pattern extraction from the big biological records. The volume of biological data is increasing quickly in daily life. The data should be in understandable format for carrying out analysis [1]. An important role has been played by the interaction between data mining and bioinformatics in the diagnosis of several diseases. 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 [2]. 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. 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 [3]. 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 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. Therefore, data mining provides efficiently predicts the sugar level within blood. In general, different studies have used data-mining methods for predicting blood sugar blood levels with and without fasting. However, some studies have tried to use data-mining approaches for predicting or classifying the postprandial blood sugar as regular or irregular. In addition, available researches on blood sugar levels in diabetic patients rely on a constant glucose screening system [4]. In this approach, a device is placed on the patient to monitor the blood sugar level of patient in specific period of time.

The general steps in diabetes predictions are described below:

- a) Data preprocessing: This 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 preprocessing approach efficiently resolves these issues [5]. 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 preprocessing prior to analyzing data [6].

b) Data Reduction: In this step, 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.

c) Classification: 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 as low, 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. The testing of classification models is carried out by comparing the forecasted values to recognized target values in a test dataset.

d) Performance Evaluation: In this step, 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. Decision tree is a fundamental classification and regression algorithm. Decision tree model has a tree like configuration. This configuration explains the procedure of classification instances on the basis of attributes [7]. 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. Neural network is a mathematical model. This model emulates the behavior of animals. This model is based on the complexity of the system for processing information. Support vector machines (SVMs) are known as supervised machine learning models. This algorithm can be used for both regression as well as classification. A Support Vector Machine (SVM) is a discriminative classification model.

II. LITERATURE REVIEW

Deeraj Shetty, 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 [8]. 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 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.

Santosh Rani, et.al (2018) stated that massive volume of clinical data was being generated at different levels of medical domain [9]. The processing and extracting this data for analysis was a very complex task because of the high volume of this data. However, it was possible to process this data using machine learning algorithms. These algorithms provided valuable information for the good treatment of diabetic patients. This information could be used for the prediction of future maladies as well. The probability of various health related issues could be predicted using the case history of patients on the basis of several parameters. Also, it was possible to develop early warning model using data mining based on association clustering and time Series. This prediction model could describe the disease along with the analysis of available metrics. The life span of the patient could be increased by providing good treatment in early stage.

Vrushali R. Balpande, et.al (2017) stated that an important role had been played by data mining in medical domain for predicting diabetes disease [10]. Diabetes was a severe health issue all over the world. According to a survey, several millions people globally were affected by this dangerous disease. Diabetes was a metabolic malady. In this disease, the fluctuation in the blood sugar level hindered the working of important body organs. This disease could be the reason of heart stroke, kidney failure, and cataract and so on. The researchers had recommended large numbers of algorithms for predicting this disease and accuracy evaluation. However, no algorithm predicted the risk level of this disease in terms of percentage to infer the effect of this disease on different body parts. The recommended model computed the risk level of this disease on different body organs.

Girdhar Gopal Ladha, et.al (2018) recognized diabetes disease as a global health issue. The researchers all over the world carried out different researches by considering various parameters [11]. The main aim of almost all research works was to detect diabetes disease in early stage. This work was focused to discover and explain a practical technique for predicting diabetes disease on the basis of available dataset. This research work provided a direction for finding different loop holes. This work could be used to develop an efficient diabetes prediction model in nearby future. This computation analysis also provided the parametric study along with the information of the features and the technique to implement the classification model.

Messan Komi, et.al (2017) stated that improper rise in the level of blood sugar was the main cause of diabetes disease [12]. This disease could be diagnosed using several conventional techniques. These techniques made use of physical and chemical tests for disease diagnosis. The techniques particularly based on the data mining could be

efficiently implemented to predict the risk level of high blood sugar. In this work, five data mining algorithms were used for predicting diabetes disease in early stage. The tested results demonstrated that ANN (Artificial Neural Network) approach outperformed the other available approaches in diabetes prediction. This approach showed highest prediction accuracy among all implemented algorithms.

Zhongxian Xu, et.al (2019) suggested the use of a risk prediction model for type 2 diabetes prediction [13]. 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. Moreover, these classifiers showed more accurate prediction results as compared to other existing classifiers. UCI Pima Indian diabetes dataset was used in this work for verifying the outcomes. 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. There were various signs that indicated the occurrence of this severe disease [14].

On the other hand, the diabetes prediction techniques used dissimilar attributes and generated different accuracy rates. 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.

Wenqian Chen, et.al (2017) stated that the growth in the popularity of data mining algorithms was noticed in healthcare domain because of their classification potential [15]. A hybrid prediction model was recommended in this work for diagnosing Type II diabetes. The recommended model made use of K-means clustering algorithm for reducing data. Also, a classification model called J48 was used for carrying out classification. Pima Indians Diabetes Dataset from UCI Machine Learning Repository was used in this work for getting the test results. The achieved results depicted that the recommended model outperformed the other existing diabetes prediction techniques in terms of accuracy. These outcomes confirmed the efficiency of recommended model in the detection of second type of diabetes disease.

Table 1: Table of Comparison

Authors Names	Year	Description	Outcomes	Research Gap
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.	The large number of features are extracted which increase execution time.
Santosh Rani, Sandeep Kautish	2018	The probability of various health related issues could be predicted using the case history of patients on the basis of several parameters.	The life span of the patient could be increased by providing good treatment in early stage.	The accuracy of the model is low which can be increased.
Vrushali R. Balpande, Rakhi D. Wajgi	2017	The researchers had recommended large numbers of algorithms for predicting this disease and accuracy evaluation.	The recommended model computed the risk level of this disease on different body organs.	The model can be tested over the real time data.
Girdhar Gopal Ladha, Ravi Kumar Singh Pippal	2018	The main aim of almost all research works was to detect diabetes disease in early stage. This work was focused to discover and explain a practical technique for predicting diabetes disease on the basis of available dataset.	This research work provided a direction for finding different loop holes.	The model can be improved which can be increase the precision and recall values.
Messan Komi, Jun Li, Yongxin Zhai, Xianguo Zhang	2017	In this work, five data mining algorithms were used for predicting diabetes disease in early stage.	This approach showed highest prediction accuracy among all implemented algorithms.	The supervised model can be applied for the comparison of diabetic prediction.
Zhongxian Xu, Zhiliang Wang	2019	The recommended model used RF-WFS and XGBoost (extreme	These results established the efficacy of the recommended	The research work can be further improved to design

		gradient boosting) classifiers for selecting best features.	model in the early detection of diabetes disease.	recommendation system.
B.V. Baiju, D. John Aravindhar	2019	Diabetic prediction based on a DIM (Disease Influence Measure) was recommended in this work for predicting diabetes.	These algorithms were compared on the basis of their performance in diabetes forecasting.	The diabetic prediction model can be improved which can increase accuracy.
Wenqian Chen, Shuyu Chen, Hancui Zhang, Tianshu Wu	2017	A hybrid prediction model was recommended in this work for diagnosing Type II diabetes.	These outcomes confirmed the efficiency of recommended model in the detection of second type of diabetes disease.	The soft voting scheme model can be designed which can improve certain parameters like precision recall and accuracy.

III. CONCLUSION

It is concluded that diabetic prediction is the approach which can predict future possibilities based on the current information. The various machine learning techniques are proposed for the diabetic prediction. The machine learning techniques are broadly classified into supervised and unsupervised learning techniques. In this paper, various supervised machine learning techniques are reviewed in terms of certain parameters.

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 Tsav, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, Ioanna Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal, **104-116, 2017**.

[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, KES, Marseille, France, **2017**.

[4] Roxana Mirshahvalad, Nastaran Asadi Zanjani, "Diabetes Prediction Using Ensemble Perceptron Algorithm", 2017, 9th International Conference on Computational Intelligence and Communication Networks

[5] Miss. Sneha Joshi, Prof. Megha Borse, "Detection and Prediction of Diabetes Mellitus Using Back-Propagation Neural Network", International Conference on Micro-Electronics and Telecommunication Engineering, **2016**.

[6] Deepa Sisodiaa, Dilip Singh Sisodiab, " Prediction of Diabetes using Classification Algorithms ", International Conference on Computational Intelligence and Data Science (ICCID), **2018**.

[7] Nongyao Nai-aruna, Rungruttikarn Moungma, "Comparison of Classifiers for the Risk of Diabetes Prediction", 7th International Conference on Advances in Information Technology, **2015**.

[8] Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, "Diabetes disease prediction using data mining", International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), **2017**.

[9] Santosh Rani, Sandeep Kautish, "Association Clustering and Time Series Based Data Mining in Continuous Data for Diabetes Prediction", Second International Conference on Intelligent Computing and Control Systems (ICICCS) **2018**.

[10] Vrushali R. Balpande, Rakhi D. Wajgi, "Prediction and severity estimation of diabetes using data mining technique", International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), **2017**.

[11] Girdhar Gopal Ladha, Ravi Kumar Singh Pippal, "A computation analysis to predict diabetes based on data mining: A review", 3rd International Conference on Communication and Electronics Systems (ICCES), **2018**.

[12] Messan Komi, Jun Li, Yongxin Zhai, Xianguo Zhang, "Application of data mining methods in diabetes prediction", 2nd International Conference on Image, Vision and Computing (ICIVC), **2017**.

[13] Zhongxian Xu, Zhiliang Wang, "A Risk Prediction Model for Type 2 Diabetes Based on Weighted Feature Selection of Random Forest and XGBoost Ensemble Classifier", Eleventh International Conference on Advanced Computational Intelligence (ICACI), **2019**.

[14] B.V. Baiju, D. John Aravindhar, "Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining", 1st International Conference on Innovations in Information and Communication Technology (ICIICT), **2019**.

[15] Wenqian Chen, Shuyu Chen, Hancui Zhang, Tianshu Wu, "A hybrid prediction model for type 2 diabetes using K-means and decision tree", 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), **2017**.