International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228139

Comparative Study of Various Data Mining Techniques for Early Prediction of Diabetes Disease

Santosh P. Shrikhande*, Prashant P. Agnihotri
School of Technology, S.R.T.M. University, Sub-Campus, Latur, Maharashtra, India

Article Info
Volume 8, Issue 1
Page Number : 287-295

Publication Issue :
January-February-2022

Article History
Accepted : 20 Feb 2022
Published: 28 Feb 2022

ABSTRACT

Diabetes is one of the prevalent diseases in the word with a high mortality rate. This disease has created several health problems and side effects on other organs of the human body. Therefore, diagnosis of this disease at early stage is essential that can reduce the fatal rate of humans. There are several ways to diagnose the diabetes but early diagnosis is quite challenging task for the medical practitioners. Recently, data mining based techniques are widely used for early prediction of diabetes that gives promising results in diabetes prediction. This paper presents the detailed review of existing data mining techniques used for diabetes prediction with their comparative study. This study also provides analysis of existing methodologies that will help in future perspective for designing and developing novel diabetes predictive models.

Keywords : Diabetes Mellitus, Diabetes Prediction, Data Mining Techniques, Machine Learning Classification Techniques

I. INTRODUCTION

Diabetes Mellitus (DM) is generally called as diabetes and it is the condition in which human body does not properly process the food for making energy [1]. When, human eat food then it is converted into glucose for creating energy. The cells of human body acquire the glucose using insulin generated from pancreas. When a human body is affected with diabetes then enough insulin is not produced or utilized for making energy for human body [1, 2]. The common symptoms of diabetes are frequent urination, weight loss, increase in thirst, increase in hunger, slow healing of wounds and giddiness etc. Diabetes effects on the other part of the human body and causes other health complications such as heart diseases, blindness, kidney failures [2, 3]. The following are the types of diabetes. The Type–I diabetes is known as Insulin Dependent Diabetes Mellitus (IDDM) or juvenile-onset diabetes. This type of diabetes is caused due to the autoimmune, genetic and environmental factors and mostly occurred in to young people those are below 30-year age. Due to this type of diabetes, beta cells of pancreas are destroyed those are responsible for creation of insulin in the body. Type II diabetes is a Non-Insulin Dependent Diabetes Mellitus (NIDDM). In this diabetes, pancreas produces insulin but in small extent which is not enough for the body's need. This type of diabetes caused due to the older age, obesity, physical

inactivity, impaired glucose tolerance, family history of diabetes, prior history of gestational diabetes etc. Type III diabetes is called as gestational diabetes that occurs in pregnant women without a previous history of diabetes. Patients of this type can control their diabetes with regular exercise and proper diet but some patients need to take medicine to control it. This diabetes cures after the pregnancy but in few cases, it may lead to the Type –II diabetes in future. The congenital diabetes is the fourth type of diabetes that occurs in the human due to genetic defects of insulin secretion, cystic fibrosis related diabetes and high doses of glucocorticoids leads to steroid diabetes [2, 3, 4]. If diabetic persons from any of above type are not undergone proper treatments then it may cause problems such as heart attacks, strokes, blindness, kidney failures and blood vessels diseases. Therefore, early diagnosis of diabetes is essential because many diabetes cases are severe due to the late diagnosis and treatment [5, 6]. Recently, data mining based techniques are widely used in early prediction of diabetes from the diabetes database. This research paper presents the systematic review of existing data mining techniques of diabetes prediction. It also illustrates the comparative study of different data mining techniques with their limitations that will help researchers for devising novel and best diabetes prediction technique. The organization of this paper is as follows. Section II, describes the design of diabetes prediction model using data mining techniques. Section III, presents literature review of different data mining techniques used for diabetes prediction. The interpretations and discussion based on the comparative study and analysis are illustrated in the Section IV. The conclusions and suggestions are provided in Section V, those will help in deciding the best technique for diabetes prediction.

II. DIABETES PREDICTION MODEL USING DATA MINING TECHNIQUES

Machine learning methods with data mining tools and techniques in early prediction of diabetes are contributing a big share in healthcare that benefits to medical practitioner for better accuracy in decision-making [4]. For this, several data mining techniques have been proposed by the researchers for early prediction of diabetes from the diabetes dataset. Following figure shows the framework of diabetes disease prediction using classification task of data mining.

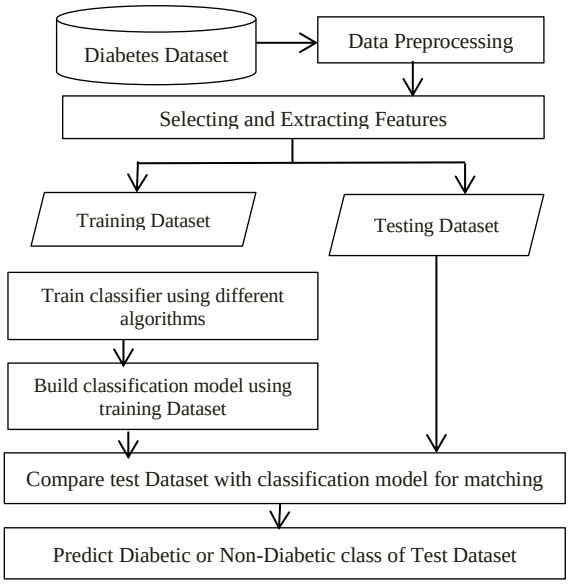

Figure 1. Diabetes Prediction Model using Data Mining Technique

The classification task of data mining is generally used for the prediction of diabetes using past data of diabetes patient's dataset [4, 7]. Classification is a supervised learning task used for predicting given dataset into predefined class labels based on the model generated during learning phase. In first phase, pre-processed data from diabetes database as a training data is feed to the predictor. Predictor learns from the training data with features and generates a model that will be further used for prediction of test data in the second phase [7, 11].

III.LITERATURE REVIEW

Data mining based techniques have been widely used in the field of medical sciences especially for the purpose of diabetes prediction and its analysis. Therefore, many researchers are developing novel approaches using various data mining and machine learning techniques for early prediction of the diabetes disease. The detailed literature survey of existing data mining techniques is given below.

Luís Chaves et al. [5], presented a comparative study of Naive Bayes, Neural Network, AdaBoost, KNN, Random Forest and SVM based classification techniques for diabetes early diagnosis. Experiment conducted using public data set of 520 instances with 17 attributes have shown the 98.08% prediction accuracy using Neural Networks.

Jobeda Jamal Khanam et al. [6], developed a diabetes prediction model using Random Forest (RF), Naïve Bayes (NB), SVM, ANN and Logistic Regression (LR). Experiments conducted on Pima Indian dataset shows the 88.6% prediction accuracy using ANN, which is better as compare to other algorithms.

Neha et al. [7], proposed a diabetes prediction system using Random Forest, Decision Tree, Naïve Bayes, SVM, KNN, and Logistic Regression (LR) and applied on Pima Indian dataset and their own collected dataset of 952 people through questionnaire. Experimental results show that random forest has given better prediction accuracy of 94.10% than others.

Md. Maniruzzaman et al. [8], designed a method for diabetes prediction using Naïve Bayes, Decision tree, Adaboost, and Random forest with K2, K5, and K10 partition protocols and applied on the National Health and Nutrition Examination Survey (NHANES) dataset of US. Experimental results show the 94.25% accuracy using random forest with K10 protocol which is better as compare to other classification algorithms.

Changsheng Zhua et al. [9], proposed a model using data mining based techniques for diabetes prediction on Pima Indian diabetes dataset. In this model, author used PCA for dimensionality reduction, K-means algorithm for clustering and logistic regression for the classification. Experimental results have shown the 97.40% diabetes prediction accuracy.

Sneha et al. [10], designed a modified approach using Decision Tree, Naïve Bayesian, SVM, Random Forest and KNN classifiers with appropriate attribute selection. Experimental results on UCI machine repository database have shown the 77% and 82.30% prediction accuracy using modified SVM and Naïve Bayesian classifier method respectively.

Desmond Bala Bisandu et al. [11], implemented the diabetes prediction system using Naïve Bayesian classifier in JAVA Language, Weka tool and MySQL. Experiments conducted on Fudawa health centre dataset have shown 95% diabetes forecasting accuracy.

Protap kumar et al. [12], applied Neural Network, SVM and Random Forest algorithms in several ways on Pima Indian dataset. Afterword, author applied different pre-processing techniques to identify the diabetes. Finally, author concluded that Neural Network has given the 80.40% prediction accuracy, which is better in comparison of all other techniques.

Souad Larabi-Marie-Sainte et al. [13], surveyed different machine learning and deep learning based techniques and implemented those not used machine learning classifiers on Pima Indian dataset to analyse their performance. The classifiers have achieved an accuracy of 68%–74%. Author recommended using these classifiers in diabetes prediction and enhancing them by developing combined models.

Sara Aboalnaser et al. [14], developed a method using KNN, Naïve Bayes, ANN, Decision Tree, Random Forest, SVM and Logistic Regression for diabetes prediction. Experiments conducted on Pima Indian dataset reveal the 99.0% diabetes prediction accuracy.

Fikirte Girma et al. [15], tested the diabetes prediction accuracy using Neural Network Back Propagation model, J48, Naïve Bayes and SVM based methods on Pima Indian dataset using R programming language. Experimental results have shown the 83.11%, 78.26%, 78.97% and 81.69% accuracy using Neural Network

Back Propagation, J48, Naïve Bayes and SVM algorithms respectively.

Yukai Li et al. [16], developed model using machine learning method which combines the feature selection and imbalanced process (SMOTE algorithm) for the diabetes prediction on the diabetes dataset of new urban area of Urumqi, Xinjiang. Afterwards author used SVM, Decision Tree and integrated learning model (Adaboost and Bagging) techniques for the prediction of diabetes. Experimental results reveals that the Adaboost algorithm has given better classification results with 94.65% G-mean and 0.9817 as area under the ROC Curve (AUC).

Dipti Sisodia et al. [17], proposed a method using Naïve Bayes, Decision Tree and SVM for diabetes prediction. Experiments conducted on Pima Indian dataset reveal the 76.30% prediction accuracy which is better as compare to the other methods.

Ashok kumar et. al. [18], developed a framework using classification tree, SVM, Logistic Regression, Naïve Bayes and ANN based computational intelligence techniques for diabetes prediction. Experimental results have shown the classification accuracy of 77% and 78% by ANN and Logistic Regression respectively.

Deepika Verma et al. [19], proposed a method for predicting breast cancer and diabetes using Naïve Bayes, SMO, REP Tree, J48 and MLP algorithms with WEKA classification tool. Experimental results have shown the 76.80% diabetes prediction accuracy using SMO algorithm on Pima Indian diabetes dataset.

Messan komi et al. [20], developed a diabetes prediction system using Gaussian mixture model (GMM), SVM, Logistic Regression (LR), Extreme Learning Machine (ELM) and ANN techniques. Experimental results have shown the 81%, 74%, 64%, 82% and 89% classification accuracy using GMM, SVM, LR, ELM and ANN.

Aparimita Swain et al. [21], proposed a method for diabetes prediction using ANN and Hybrid Adaptive Neuro-Fuzzy Inference System. The ANN was trained using 100 individuals with mean age of 42 years with an equal proportion of male and female dataset in MATLAB. Experimental results reveal that the ANFIS approach is more acceptable as compare to ANN approach in terms of accuracy.

Sajida Perveen et al. [22], proposed a model using adaptive boosting and bagging ensemble techniques using J48 (c4.5) and standalone J48 for diabetic classification. Author used this method for three different ordinal adult groups in Canadian Primary Care Sentinel Surveillance Network (CPCSSN) database which contains 667907 patient records. Experimental results have shown that, overall performance of AdaBoost ensemble method is better than bagging as well as standalone J48 decision tree.

Sadri et al. [23], developed a method using Naïve Bayes, RBF Neural Network and J-48 algorithms for diabetes prediction in Weka software. Experiments conducted on Pima Indian dataset reveal the 76.95% prediction accuracy using Naïve Bayes which is better as compare to other algorithms.

Aruna Pavate et al. [24], developed a system using genetic algorithm, nearest neighbour and fuzzy rule-based system to predict diabetes and its complications. The dataset of 235 patients were collected using a questionnaire reviewed by health experts. The performance of diabetes prediction system is tested on subset of features generated by this implemented algorithm. The genetic algorithm version 3 has given 95.50% diabetes prediction accuracy.

Gaganjot Kaur et al. [25], proposed a new approach using modified J48 algorithm for efficiently prediction of the diabetes from Pima Indian dataset. Experimental results have shown the 99.87% prediction accuracy, which is significant improvement over the existing J-48 algorithm.

IV. COMPARATIVE STUDY AND DISCUSSION

This section presents the details of diabetes disease prediction accuracy obtained by the previous researchers with complete analysis of what type of data set and methodologies were used by selecting

important features for conducting experiments. Following table Table 1 shows the diabetes prediction accuracy achieved by different researcher using machine learning classification techniques.

TABLE I
DIABETES PREDICTION ACCURACY OF EXISTING DATA MINING TECHNIQUES

Ref. No.	Year	Technique Used	Prediction Results	Dataset Used
Luís Chaves et al. [5]	2021	Neural Network(NN), KNN, Support Vector Machine(SVM), AdaBoost, Random Forest(RF), Naïve Bayes (NB)	NN=98.08%, KNN=97.31% SVM=97.12%, AdaBoost=97.30%, RF=96.90%, NB=86.92%	Publically available dataset of Sylhet Diabetes Hospital in Sylhet, Bangladesh
Jobeda Jamal et al.[6]	2021	ANN, Random Forest (RF), Naïve Bayes(NB), SVM, and Logistic Regression (LR)	ANN=88.6%, RF=77.34%, NB=78.28%, SVM=78.0%, LR=87.85%	Pima Indian Diabetes Dataset
Neha et al.[7]	2020	Random Forest (RF), Decision Tree(DT), Naïve Bayes(NB), SVM, KNN, and Logistic Regression (LR)	RF=94.10%, DT=84.00%, NB=80.6%, SVM=86.5%, KNN=77.3%. LR=85.7%	Dataset of 952 patients collected through questionnaire
Md. Maniruzzaman et al. [8]	2020	Naïve Bayes (NB), Decision tree (DT), Adaboost (AB), Random forest (RF) with K2, K5, and K10 partition protocols	NB=86.70%, DT=89.65%, AB=92.93%, RF =94.25% with K10 partition protocol	National Health and Nutrition Examination Survey (NHANES) dataset of US
Changsheng Zhu et al. [9]	2019	PCA + K-means + Logistic Regression	PCA+K-means+Logistic Regression=97.40%, PCA+K-means = 79.94%	Pima Indian Diabetes Dataset
N. Sneha et al. [10]	2019	Modified approach using Decision Tree, Naïve Bayesian (NB), SVM, Random Forest and KNN with appropriate attribute selection.	Prediction accuracy using modified SVM=77%, NB=82.30%	Pima Indian Diabetes Dataset
Desmond Bala et al. [11]	2019	Naïve Bayes Classifier	Prediction accuracy of Naïve Bayes = 95%	Dataset of 155 real diabetes patients.
Protap Kumar et al. [12]	2019	Artificial Neural Network(ANN), SVM, Random Forest(RF)	ANN=80.40%, SVM=75.4%, RF=77%	Pima Indian Diabetes Dataset
Souad Larabi-Marie-Sainte et al. [13]	2019	REP Tree, KStar, OneR, PART, SMO and BayesNet	REP Tree= 74.48%, KStar= 68.23%, OneR=70.83%, PART=74.35, SMO= 72.14%, BayesNet= 73.83%	Pima Indian Diabetes Dataset
Sara Aboalnaser et al. [14]	2018	KNN, Naïve Bayes (NB), ANN, Decision Tree, Random Forest, SVM and Logistic Regression.	KNN=99.0%, NB=76.40%, ANN=80.10%,DT=93.60%, RF=96.90%, SVM =95.80% LR=77.60%.	Pima Indian Diabetes

				Dataset
Fikirte Girma et al. [15]	2018	Back Propagation algorithm, J48, Naïve Bayes, SVM	Back Propagation =83.11%, J48=78.26, NB=78.97%, SVM=81.69%	Pima Indian Diabetes Dataset
Yukai Li et al. [16]	2018	Combined feature selection and imbalanced process (SMOTE algorithm), SVM, Decision tree and integrated learning model (Adaboost and Bagging) techniques.	Adaboost = 94.84%, SVM= 92.62%, Decision tree = 91.15%, Integrated learning model (Adaboost and Bagging) = 91.15%	Diabetic patient health management data of New Urban Area of Urumqi, Xinjiang
Deepti Sisodiaa et al. [17]	2018	Decision Tree, SVM and Naive Bayes algorithms	Prediction accuracy using Naive Bayes = 76.30%	Pima Indian Diabetes Dataset
Ashok Kumar et. al. [18]	2017	Logistic Regression (LR), ANN, SVM, KNN, Naïve Bayes (NB) and Classification Tree (CT).	LR=78%, ANN=77%, SVM=74%, KNN=73%, NB=75%, CT = 70%	Pima Indian Diabetes Dataset
Deepika Verma et. al. [19]	2017	SMO, Naïve Bayes, REP Tree, J48 and MLP algorithms in WEKA classification tool	SMO = 76.80%, Naïve Bayes = 75.75%, REP Tree = 74.46%, J48 = 74.4% and MLP = 74.75%	Pima Indian Diabetes Dataset
Messan Komi et al.[20]	2017	Gaussian mixture model (GMM), SVM, Logistic regression (LR), Extreme Learning Machine (ELM) and ANN techniques.	Accuracy of GMM=81%, SVM=74%, LR=64%, ELM=82% and ANN =89%	Clinical Diabetes Database
Aparimita Swain et.al. [21]	2016	ANN and Hybrid Adaptive Neuro-Fuzzy Inference System	Accuracy using ANFIS = 90.32 % and ANN = 71.10 %	Pima Indian Diabetes Dataset
Sajida Perveen et al. [22]	2016	Adaptive Boosting and Bagging ensemble techniques using J48 (c4.5) decision tree along with standalone J48 technique	Area under AROC for bagging ensemble is 0.98. Performance of Adaptive Boosting ensemble is better than other methods	Canadian Primary Care Sentinel Surveillance Network (CPCSSN) Database
Sadri Sa'di et al. [23]	2015	Naive Bayes, RBF Network, and J48 classifiers	Naive Bayes=76.95%, RBF Network=74.34%, and J48=76.52%	Pima Indian Diabetes Dataset
Aruna Pavate et. al. [24]	2015	Genetic Algorithm (GA), Nearest Neighbour and Fuzzy Rule-Based System	Genetic Algorithm Version 3 = 95.50% prediction accuracy	235 Patient's data collected using health experts questionnaire
Gaganjot Kaur et al. [25]	2014	Improved J48 Classification Algorithm	Improved J48 = 99.87%, J48 = 73.82%	Pima Indian Diabetes Dataset

After exhaustive study of diabetes prediction using different approaches of data mining techniques, it has been observed that most of existing methodologies are based on the classification and clustering tasks of data mining. Some machine learning methods with data mining tools and techniques have been mostly used for diabetes prediction. Moreover, some methodologies have been adopted ensemble techniques by combining more than one algorithm together such as Adaptive Boosting and Bagging, PCA, ANFIS and ANN, PCA, K-means and Logistic Regression for the diabetes prediction. These methodologies designed using machine learning, data mining tools and techniques have shown the acceptable diabetes predictions results, but with limitations in some factors such as diabetes dataset and methodologies adopted. From the study, it has been observed that the Pima Indian dataset with only nine attributes from UCI machine repository was widely used for conducting experiments. Very few researchers have used primary dataset of diabetes such as clinical dataset or own collected data from patients using questionnaires designed by health experts. The Pima Indian dataset was collected from University of California, Irvine (UCI) machine learning repository consisting records of 768 patients with only 9 attributes of Pima Indian population living in Arizona, USA. It is reported that most of the patients from this database were diagnosed diabetic due to the obesity. It is found better prediction accuracy using Pima Indian dataset focusing on obesity related attributes as compare to other primary datasets. If dataset of diabetes patients from other country or region is used for conducting an experiment then same results and conclusion cannot be drawn. Therefore, if we want to validate the real performance and robustness of any diabetes prediction algorithm then more number of databases from different regions and/or countries should be focused. Moreover, more number of appropriate attributes along with the pre-processing techniques needs to be used rather than only nine attributes from Pima Indian dataset. Some of the authors have reported good and acceptable diabetes prediction results using Artificial Neural Network (ANN) based methods. Some of the researchers have designed their methodologies using only one single technique and some researchers have used ensemble techniques to verify the performance of diabetes prediction. It has been observed that the methodologies using ensemble techniques have improved their prediction performance in some extent than individual technique. Therefore, the methodologies with best ensemble technique along with pre-processing of input dataset, appropriate attributes selection and their representation plays an important role in designing the robust and best approach for the diabetes disease prediction.

V. CONCLUSIONS

This research paper has presented a detailed review and comparative study of existing data mining techniques used in early prediction of the diabetes disease. After exploring and reviewing various data mining techniques used in diabetes prediction along with their experimental results following major observations are highlighted. The diabetes prediction accuracy of any model is mainly dependent on the dataset and algorithm used. Appropriate attributes selection and representation also plays a vital role in improving performance of diabetes prediction system. Pima Indian diabetes dataset with only nine attributes has been mostly used to verify the performance of diabetes prediction. Some diabetes prediction models have worked very well using Pima Indian dataset but same performance cannot be achieved using other diabetes dataset. Diabetes prediction system using ensemble techniques have shown improvement in some extent as compare to one single technique. Therefore, it is recommended to focus on different clinical datasets of diabetes patient with more number of appropriate attributes to validate the real performance of diabetes prediction system because prediction accuracy in medical field is most important. An appropriate pre-processing technique for attribute selection and hybrid or ensemble methods for classification should be focused on to devise novel diabetes prediction system. Moreover, some advanced classifiers along with the machine learning, deep learning and genetic algorithms based techniques can be focused to design best diabetes prediction system. Our future work is to focus on these addressed issues and design a novel methodology for early prediction of diabetes disease.

VI. REFERENCES

[1]. American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus", Diabetes Care, Volume 37, Supplement 1, pp. s81-s90, January 2014.

[2]. Zidian Xie, Olga Nikolayeva, Jiebo Luo, "Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques", Preventing Chronic Disease, Public Health Research, Practice, and Policy, Vol. 16, E130, pp. 1-9, September - 2019.

[3]. Madhusmita Rout, Amandeep Kaur, "Prediction of Diabetes Risk based on Machine Learning Techniques", International Conference on Intelligent Engineering and Management (ICIEM –IEEE), pp. 246-251, June-2020.

[4]. Surabhi Kaul, Yogesh Kumar, "Artifcial Intelligence-based Learning Techniques for Diabetes Prediction: Challenges and Systematic Review", SN Computer Science, Springer Nature Journal, pp. 1-7, October 2020.

[5]. Luís Chaves and Gonçalo Marques, "Data Mining Techniques for early Diagnosis of Diabetes: A Comparative Study", Applied Sciences (MDPI) 11, 2218, pp.1-12, March-2021.

[6]. Jobeda J. Khanam, Simon Y. Foo, "A comparison of machine learning algorithms for diabetes prediction", ICT Express, published by Elsevier, https://doi.org/10.1016/j.icte.2021.02.004, pp. 1-8, 2021.

[7]. Neha Prerna Tiggaa, Shruti Garga, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods", International Conference on Computational Intelligence and Data Science, Procedia Computer Science, pp. 706–716, 2020.

[8]. Md. Maniruzzaman, Md. Jahanur Rahman, Benojir Ahammed, Md. Menhazul Abedin, "Classifcation and prediction of diabetes disease using machine learning paradigm", Health Information Science and Systems (Springer Nature), https://doi.org/10.1007/s13755-019-0095-z, pp. 1-14, 2020.

[9]. Changsheng Zhu, Christian Uwa Idemudia and Wenfang Feng, "Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques", Informatics in Medicine Unlocked, https://doi.org/10.1016/j.imu.2019.100179, pp. 1-7, 2019.

[10]. N. Sneha, Tarun Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection", Journal of Big Data, Springer Open Journal, pp.1-19, 2019.

[11]. Desmond Bala Bisandu, Dorcas Dachollom Datiri, Eva Onokpasa, Godwin Thomas, Musa Maaji Haruna, Aminu Aliyu, Jerry Zachariah Yakubu, "Diabetes Prediction Using Data Mining Techniques", International Journal of Research and Innovation in Applied Science (IJRIAS),Volume IV, Issue VI, pp. 103-111, June-2019.

[12]. Protap Kumar Saha, Nazmus Sakib Patwary, Ifthakhar Ahmed, "A Widespread Study of Diabetes Prediction Using Several Machine Learning Techniques", 22nd International Conference on Computer and Information Technology (ICCIT-IEEE), pp.1-5, December 2019.

[13]. Souad Larabi-Marie-Sainte, Linah Aburahmah, Rana Almohaini and Tanzila Saba, "Current Techniques for Diabetes Prediction: Review and Case Study", MDPI Applied Sciences Journal, pp. 1-18, October-2019.

[14]. Sara A. Aboalnaser, Hanan R. Almohammadi, Comprehensive Study of Diabetes Miletus Prediction using Different Classification Algorithms, IEEE Conference on Developments in eSystems Engineering (DeSE), pp.128-133, 2019.

[15]. Fikirte Girima Woldemichael, Sumitra Menaria, "Prediction of Diabetes using Data Mining Techniques", Proceeding of 2nd International Conference on Trends in Electronics and Informatics (ICOEI- IEEE Xplorer), pp. 414-418, May-2018.

[16]. Yukai Li, Huling Li, and Hua Yao, "Analysis and Study of Diabetes Follow-Up Data Using a Data-Mining-Based Approach in New Urban Area of Urumqi, Xinjiang, China, 2016-2017", Hindawi, Computational and Mathematical Methods in Medicine, Volume 2018, Article ID-7207151, pp.1-8, July-2018.

[17]. Deepti Sisodiaa, Dilip Singh Sisodiab, "Prediction of Diabetes using Classification Algorithms", International Conference on Computational Intelligence and Data Science, Elsevier Procedia Computer Science, pp. 1578-1585, 2018.

[18]. Ashok Kumar Dwivedi, "Analysis of computational intelligence techniques for diabetes mellitus prediction", Springer's Neural Computing Applications, pp. 1-9, April-2017.

[19]. Deepika Verma, Nidhi Mishra, "Analysis and Prediction of Breast cancer and Diabetes disease datasets using Data mining classification Techniques", Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS-IEEE Xplorer), pp. 533-538, December-2017.

[20]. Messan Komi, J un Li, Y ongxin Zhai, Xianguo Zhang, "Application of Data Mining Methods in Diabetes Prediction", IEEE, 2nd International Conference on Image, Vision and Computing, pp.1006 -1010, June-2017.

[21]. Aparimita Swain, Sachi Nandan Mohanty, Ananta Chandra Das, "Comparative Risk Analysis on Prediction of Diabetes Mellitus using Machine Learning Approach", International Conference on Electrical, Electronics and Optimization Techniques (ICEEOT-IEEE Xplorer), pp. 3312-3317, March-2016.

[22]. Sajida Perveena, Muhammad Shahbaza, Aziz Guergachib, Karim Keshavjeec, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes", Elsevier Procedia Computer Science of Symposium on Data Mining Applications, pp. 115-121, March 2016.

[23]. Sadri Sadi, Amanj Maleki, Ramin Hashemi, Zahra Panbechi, Kamal Chalabi, "Comparision of Data Mining Algorithms in Diagnosis of Type II Diabetes", International Journal on Computational Science and Applications (IJCSA) Vol.5, No.5, pp. 1-12, October 2015.

[24]. Aruna Pavate, Nazneen Ansari, "Risk Prediction of Disease Complications in Type-2 Diabetes Patients Using Soft Computing Techniques", IEEE Fifth International Conference on Advances in Computing and Communications, pp.371-375, 2015.

[25]. Gaganjot Kaur, Amit Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes", International Journal of Computer Applications, Vol. 98, no.22, pp.13-17, July-2014.

[26]. Rakesh Motka, Viral Parmar Balbindra Kumar, A. R. Verma, "Diabetes Mellitus Forecast Using Different Data Mining Techniques", IEEE 4th International Conference on Computer and Communication Technology (ICCCT), pp. 99-103, September-2013.

[27]. Vrushali Balpande, Rakhi Wajgi, "Review on Prediction of Diabetes using Data Mining Technique", International Journal of Research and Scientific Innovation (IJRSI) | Volume IV, Issue IA, pp. 43-46, January 2017.

[28]. Veena Vijayan, Aswathy Ravikumar, "Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus", International Journal of Computer Applications, Volume 95 - No.17, pp. 12-16, June.

Cite this article as :

Santosh P. Shrikhande, Prashant P. Agnihotri, "Comparative Study of Various Data Mining Techniques for Early Prediction of Diabetes Disease", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 1, pp. 287-295, January-February 2022. Available at doi : https://doi.org/10.32628/CSEIT228139
Journal URL : https://ijsrcseit.com/CSEIT228139