Original Article

A survey on diabetes risk prediction using machine learning approaches

Shimoo Firdous[1], Gowher A. Wagai[2], Kalpana Sharma[1]

[1]Department of Computer Science, Bhagwant University, Ajmer, Rajasthan, [2]Department of Medicine-Associated Hospital GMC, Anantnag, Jammu and Kashmir, India

Abstract

Background: Diabetes mellitus (DM) is a chronic condition that can lead to a variety of consequences. Diabetes is a condition that is caused by factors such as age, lack of exercise, sedentary lifestyle, family history of diabetes, high blood pressure, depression and stress, poor food, and so on. Diabetics are at a higher risk of developing diseases such as heart disease, nerve damage (diabetic neuropathy), eye problems (diabetic retinopathy), kidney disease (diabetic nephropathy), stroke, and so on. According to the International Diabetes Federation, 382 million people worldwide suffer from diabetes. By 2035, this number will have risen to 592 million. Every day, a large number of people become victims, and many are ignorant whether they have it or not. It primarily affects individuals between the ages of 25 and 74 years. If diabetes is left untreated and undiagnosed, it can lead to a slew of complications. The emergence of machine learning approaches, on the other hand, solves this crucial issue. **Aims and Objectives:** The aim was to study the DM and analyze how machine learning algorithms are used to identify the diabetes mellitus at an early stage, which is one of the most serious metabolic disorders in the world today. **Methods and Materials:** Data was obtained from databases such as Pubmed, IEEE xplore, and INSPEC,and from other secondary sources and primary sources in which methods based on machine learning approaches used in healthcare to predict diabetes at an early stage are reported. **Results:** After surveying various research papers, it was found that machine learning classification algorithms like Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF) etc shows the best accuracy for predicting diabetes at an early stage. **Conclusion:** Early detection of diabetes is critical for effective therapy. Many people have no idea whether or not they have it. The full assessment of Machine learning approaches for early diabetes prediction and how to apply a variety of supervised and unsupervised machine learning algorithms to the dataset to achieve the best accuracy are addressed in this paper.. Furthermore, the work will be expanded and refined to create a more precise and general predictive model for diabetes risk prediction at an early stage. Different metrics can be used to assess performance and for accurate diabetic diagnosis.

Keywords: Accuracy, classification, diabetes mellitus, machine learning algorithm

Introduction

Diabetes mellitus is a metabolic disorder defined by abnormally high blood sugar levels due to a lack of insulin secretion or a combination of insulin resistance and insufficient insulin synthesis to compensate.[1] It is a progressive metabolic ailment that affects all parts of the patient's life, including physical and mental well-being, and no therapy technique can produce spectacular improvements or stop the disease from progressing.[2] In the year 2000, India had the greatest number of diabetics in the world (31.7 million), which increased to 62.4 million in 2011 and is anticipated to reach 69.9 million by 2025.[3,4] Rapid urbanization and economic development are to blame for India's high frequency. Indians are more likely to develop diabetes as a result of their low BMI combined with high upper-body adiposity, high body fat percentage, and high insulin resistance.[5] Blurred vision, weight loss, fatigue, increased hunger and thirst, confusion, frequent urination, poor healing, frequent

Address for correspondence: Dr. Gowher A. Wagai, Al-Jowher Diabetes Care and Scan Centre, Nazuk Mohalla, Cheeni Chowk, Anantnag - 192 101, Jammu and Kashmir, India.
E-mail: drgowherahmadwagai@yahoo.com

Received: 02-03-2022 **Revised:** 24-06-2022
Accepted: 26-06-2022 **Published:** 16-12-2022

Access this article online

Quick Response Code:

Website: www.jfmpc.com

DOI: 10.4103/jfmpc.jfmpc_502_22

How to cite this article: Firdous S, Wagai GA, Sharma K. A survey on diabetes risk prediction using machine learning approaches. J Family Med Prim Care 2022;11:6929-34.

infections, and difficulty concentrating are all signs or symptoms of diabetes. "Diabetes means you have too much sugar in your blood. High blood sugar problems start when your body no longer makes enough of a chemical, or hormone, called insulin."[6] "Sweet urine" is the direct translation. Normal urine does not contain sugar. There is sugar (or more precisely glucose) in the urine because the amount of glucose in the blood has increased to the point where it spills over into the urine. Because the body is unable to metabolize glucose properly, it accumulates in the bloodstream. As a result, diabetes is a disease in which the body is unable to use glucose properly.

This paper provides many machine learning algorithms used for the early prediction of diabetes. The remainder of the paper is conceived in the following manner: Section (1) is the introduction; section (2) is diabetes and its types; section (3) is machine learning algorithms; section (4) literature survey for prediction of diabetes; and section (5) is the conclusion.

Diabetes and its Types

Diabetes mellitus (DM) is a metabolic disease with a variety of causes. It is characterized by persistent hyperglycemia and alterations in carbohydrate, lipid, and protein metabolism caused by insulin deficiency, insulin action, or both. Diabetes is a chronic disease. Diabetes can injure neurons and blood arteries in the eyes, kidneys, heart, and lower legs if not effectively treated. Problems may emerge if blood glucose levels remain high for an extended period. Gum disease or tooth decay are examples of mouth issues. Diabetic retinopathy is a condition that causes vision loss and, in severe cases, blindness. Heart and blood vessel illnesses include heart attacks, strokes, and peripheral artery disease (cardiovascular diseases or CVD) (insufficient blood supply to the feet and legs). Kidney disease (diabetic nephropathy) is a condition in which the kidneys do not function properly or at all.[7] The three kinds of diabetes include type 1 diabetes, type 2 diabetes, and gestational diabetes.

Type 1 diabetes (T1D)

The body does not produce enough insulin in type 1 diabetes. Body cells can't absorb glucose from the bloodstream without insulin, so they have to rely on other sources of energy. An excess of glucose in the blood causes diabetes and its complications. This type of diabetes is also known as insulin-dependent diabetes mellitus (IDDM). Although it affects adolescents and teenagers more frequently, it can affect anyone at any age. It requires a delicate balancing act of insulin injections (and, in some circumstances, oral medicines), exercise, nutrition planning, and lifestyle changes. Frequent urination, unusual thirst, unusual hunger, rapid weight loss, weariness and weakness, nausea, and irritability are some of the symptoms of type 1 diabetes.

Type 2 diabetes (T2D)

Type 2 diabetes can range from primarily insulin resistance to mostly secretory dysfunction with or without insulin resistance. The pancreas produces insulin, but it may not be enough to keep blood glucose levels normal, or the cells may be resistant to the insulin produced. The illness is most prevalent in those over the age of 40, but it is also becoming more prevalent in teenagers and young children. Type 1 diabetes is characterized by drowsiness, dry, itchy skin, unintended weight gain or loss, blurred vision, tingling, numbness, pain in the lower legs, easy weariness, sluggish healing of cuts, or scratches, and frequent infections (e.g., vaginal infections). Food, activity, lifestyle control, and, in some situations, oral medicines or insulin are all necessary for type 2 diabetes.[6]

Gestational diabetes

Pregnant women who have never had diabetes before but have high blood glucose (sugar) levels during pregnancy are diagnosed with gestational diabetes. It is a temporary condition that affects 2%–4% of all pregnant women and usually disappears after the baby is born. Women who have had gestational diabetes in the past are more likely to develop type 2 diabetes later in life. There is no known etiology for this type of diabetes. The placenta supports the infant's development; placental hormones help the baby develop, but they also prevent the mother's insulin from working properly in her body, resulting in insulin resistance. When a mother's body is unable to create and use all of the insulin required during pregnancy, gestational diabetes develops. The majority of women are unaware of any signs or symptoms of gestational diabetes. Increased thirst and more frequent urination are two symptoms.

Diabetes mellitus is a deadly disorder, if not treated early' however, early detection can minimize the risk significantly. A range of medical diagnostic procedures is already in use for early diagnosis. Early risk forecasts can be made using machine learning techniques. Recent research has given promising results in terms of forecasting the risk of diabetes mellitus. Machine learning is a field of study in which algorithms are used to teach machines without the need of humans. Without having to explicitly program them, we can train them to do a given job and then use that training to handle similar duties. Accuracy is always a major problem in medical science, and different algorithms might yield varying degrees of accuracy on the same data set. To design a better classifier for better classification, it is vital to figure out which algorithm delivers the greatest results. Machine learning can now be found in almost every industry. Its application in medical science has the potential to improve healthcare dramatically.

Decision trees, random forests, support vector machines, naive Bayes classifiers, and artificial neural networks are examples of machine learning and classification algorithms that work well in risk prediction. Because of the algorithms' computing and data management skills, this is possible. Measures of classification accuracy can be used to select the best algorithm and determine the best classification accuracy. This statistic, however, is insufficient to properly and efficiently determine the best method.

When determining the best conclusion, other variables such as the receiver operating characteristic (ROC) value, F-score, and calculation time should be taken into account. Metrics include classification accuracy, F-score, ROC value, and computation time. Future researchers will be aided by the findings of this study in constructing a baseline strategy for DM classification.

Machine Learning Algorithms

Machine learning (ML) is a rapidly developing field that is being applied in a variety of medical applications. ML models all learn from the past and make predictions based on a data set. Diabetes detection will become much easier and less expensive thanks to recent advances in ML. There are numerous diabetic data sets accessible. As a result, ML is required for medical diagnostics. The goal of this study is to forecast a patient's probability of developing diabetes. Algorithms for machine learning are employed. There are two types of learning for the study.
1) Supervised Learning
2) Unsupervised Learning.

The goal of a supervised learning algorithm is to predict based on labeled data. In supervised learning, the data is labeled. It simulates what a student might learn from an instructor. Unsupervised learning, on the other hand, does not label the data. It's more like self-learning based on previous experiences. The goal is to forecast a variable's value. A set of traits and features are used to represent the data. The outcome of guided learning is predetermined. Decision trees (DT), random forests, linear regression, logistic regression, naive Bayes classifiers, k-nearest neighbors (k-NN), support vector machine (SVM), and artificial neural networks (ANN) are some of the most commonly used techniques.

The data in unsupervised learning is made up of values without labels, and the outcome is not predetermined. Based on self-learning, the model makes predictions. Forecasting, classifying, detecting, segmenting, and categorizing data are the key goals of these models. Machine learning applications include analysis, recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.

Literature Survey for Prediction of Diabetes using Machine Learning Approaches

Birjais *et al.*[8] experimented on PIMA Indian Diabetes (PID) data set. It has 768 instances and 8 attributes and is available in the UCI machine learning repository. They aimed to focus more on diabetes diagnosis, which, according to the World Health Organization (WHO) in 2014, is one of the world's fastest-growing chronic diseases. Gradient boosting, logistic regression, and naive Bayes classifiers were used to predict whether a person is diabetic or not, with gradient boosting having an accuracy of 86%, logistic regression having a 79% accuracy, and naive Bayes having a 77% accuracy.

Sadhu, A. and Jadli A.[9] experimented on a diabetes data set taken from the UCI repository. There were 520 occurrences and 16 attributes in all. They attempted to concentrate their efforts on predicting diabetes at an early stage. On the validation set of the employed data set, seven classification techniques were implemented: k-NN, logistic regression, SVM, naive Bayes, decision tree, random forests, and multilayer perceptron. The random forests classifier proved to be the best model for the concerned data set, with an accuracy score of 98%, followed by logistic regression at 93%, SVM at 94%, naive Bayes at 91%, decision tree at 94%, random forests at 98%, and multilayer perceptron at 98%, according to the results of training several machine learning models.

Xue *et al.*[10] experimented on the diabetes data set taken from the UCI repository; there were 520 patients and 17 qualities in it. They attempted to concentrate on early detection of diabetes. They trained on the actual data of 520 diabetic patients and probable diabetic patients aged 16–90 using supervised ML techniques such as SVM, naive Bayes classifiers, and LightGBM. The performance of the SVM is the best when comparing classification and recognition accuracy. The naive Bayes classifier is the most widely used classification algorithm, with an accuracy of 93.27%. SVM has the highest accuracy rate of 96.54%. LightGBM has an accuracy of only 88.46%. This demonstrates that SVM is the best classification algorithm for diabetes prediction.

Le *et al.*[11] experimented on the early-stage diabetes risk prediction; the data set used in this research was taken from the UCI repository and consisted of 520 patients and 16 variables. They suggested a ML approach for predicting diabetes patients' early onset. It was a new wrapper-based feature selection method that employed grey wolf optimizer (GWO) and adaptive particle swarm optimization (APSO) to optimize the multilayer perceptron (MLP) and reduce the number of needed input attributes. They also compared the results obtained with this method to those obtained via a variety of traditional machine learning algorithms, including SVM, DT, k-NN, naive Bayes classifier (NBC), random forest classifier (RFC), and logistic regression (LR). LR achieved a 95% accuracy rate. k-NN had a 96% accuracy rate, SVM a 95% accuracy rate, NBC a 93% accuracy rate, DT a 95% accuracy rate, and RFC had a 96% accuracy rate. The suggested methods' computational findings show that not only are fewer features required but also that higher prediction accuracy may be attained (96% for GWO–MLP and 97% for APSO–MLP). This research has the potential to be applied in clinical practice and used as a tool to assist doctors and physicians.

Julius *et al.*[12] used the Waikato Environment for Knowledge Analysis (Weka) application platform to test a data set collected from the UCI repository. There were 520 samples in the data set, each with a collection of 17 attributes. The goal of this study was to use machine learning classification approaches based on observable sample attributes to predict diabetes at an early stage.

The k-NN, SVM, functional tree (FT), and RFCs were employed as classifiers. k-NN had the highest accuracy of 98%, followed by SVM at 94%, FT at 93%, and RF at 97%.

Shafi *et al*.[13] reported that because diabetes is a serious illness, early detection is always a struggle. This study used machine learning classification methods to develop a model that could solve any problem and that could be used to identify diabetes development early on. The authors of this research made concerted efforts to develop a framework that could accurately predict the likelihood of diabetes in patients. As part of this study, the three ML approach classification algorithms—DT, SVM, and NBC—were studied and assessed on various measures. In the study, the PID data set acquired from the UCI repository was used to save time and produce precise findings. The experimental results suggested that the NBC approach was adequate, with a 74% accuracy, followed by SVM with a 63% accuracy and the DT with a 72% accuracy. In the future, the built framework, as well as the ML classifiers used, could be used to identify or diagnose other diseases. The study, as well as several other ML methodologies, could be extended and improved for diabetes research, and the scientists intended to classify other algorithms with missing data.

Khanam *et al*.[14] experimented with diabetes illness prediction. Diabetes is a condition with no known cure; therefore early detection is essential. In this study, data mining, ML techniques, and neural network (NN) methodologies were utilized to predict diabetes. They developed a technique that could accurately predict diabetes. They used data from the UCI repository's PID data set. The data set included information on 768 patients and their 9 attributes. On the data set, they utilized seven ML methods to predict diabetes: DT, k-NN, RFC, NBC, AB, LR, and SVM. They used the Weka tool to preprocess the data. They discovered that a model combining LR and SVM is effective at predicting diabetes. They created a NN model with two hidden layers and varied epochs and found that the NN with two hidden layers gave 88.6% accuracy. ANN scored 88.57%, LR scored 78.85%, NBC scored 78.28%, and RFC scored 77.34%.

Sisodia *et al*.[15] used the PID data set available on the UCI repository. This data set contained 768 patients and 8 attributes. They employed three ML classifications to identify diabetic patients: DT, SVM, and NBC. NBC had the highest accuracy (76.30%) when compared to the other models.

Agarwal *et al*.[16] used the PID data set of 738 patients as well in their study. To analyze the effectiveness of this data set for identifying diabetic patients, the authors applied models such as SVM, k-NN, NBC, ID3, C4.5, and CART. The SVM and LDA algorithms were the most accurate, with an accuracy of 88%.

Rathore *et al*.[17] employed classification techniques like SVM and DTs to predict diabetes mellitus. The PID data set provided the data for this investigation. PIMA India prioritizes women's health. The SVM has an accuracy of 82%.

To predict diabetes mellitus, Hassan *et al*.[18] employed classification approaches such as the DT, k-NN, and SVM. The SVM outperformed the DT and KNN methods with a maximum accuracy of 90.23%.

Kandhasamy and Balamurali[19] investigated the prediction accuracy of J48, k-NN, RFC, and SVM on the diabetes data set. Before preprocessing the data, the author discovered that the J48 method had a higher accuracy than others, at 73.82%. After preprocessing, k-NN and RFC demonstrated improved accuracy.

Meng *et al*.[20] examined J48, LR, and k-NN algorithms on the diabetes data set. J48 was found to be the most accurate, with a classification accuracy of 78.27%.

Nai-Arun and Moungmai[21] created a web application based on the prediction accuracy for diabetes prediction. They compared prediction methods such as DTs, NNs, LR, NBC, and RFC, as well as, bagging and boosting. They discovered that RFC performed best in terms of accuracy and ROC score, with an accuracy of 85.558% and an ROC value of 0.912.

Saravananathan and Velmurugan[22] looked at J48, CART, SVM, and k-NN on a medical data set in their research. They compared them based on accuracy, specificity, sensitivity, precision, and error rate. With a score of 67.15%, they discovered that J48 algorithms were the most accurate, followed by SVM (65.04%), CART (62.28%), and k-NN (53.39%).

Kumari and Chitra[23] used SVM, RFC, DT, MLP, and LR, as well as four k-fold cross-validations (k = 2,4,5,10) in their research. According to the researchers, MLP with four-fold cross-validation achieves the best accuracy, at 78.7%. They discovered that MLP outscored all other algorithms.

To predict diabetes, Kavakiotis *et al*.[24] employed NBC, RFC, k-NN, SVM, DT, and LR methods. The algorithms were applied using a ten-fold cross-validation technique. SVM had the best accuracy of all the approaches, measuring 84%, according to the study.

The work on the classification of "Diabetes Prediction" based on eight attributes was done by Rawat *et al*.[25] In this study, five ML algorithms for the analysis and prediction of diabetic patients were described: AdaBoost, LogicBoost, RobustBoost, naive Bayes, and bagging. A group of diabetic PIMA Indians was used to test the proposed strategies. The computed results were found to be quite accurate, with a classification accuracy of 81.77% and 79.69% for the bagging and AdaBoost techniques, respectively. As a result, the proposed DM prediction algorithms were particularly appealing, effective, and efficient.

Using disease classifiers and an actual data set, Nai-Arun and Moungmai[21] suggested a web application. The data for this component was collected from 30,122 people at Sawanpracharak Regional Hospital's twenty-six primary care units between 2012 and 2013. To identify a predictive model, thirteen classification

models were investigated before the web application was created. These models, except the RFC method, included the DT, NN, LR, NBC, and RFC algorithms, which all used a combination of bagging and boosting techniques. Each model's accuracy and ROC curves were calculated and compared to others to see how robust they were. According to the findings, RFC won in both accuracy and ROC curve. This could be owing to a wide range of options. Not only were data and input factors chosen at random in the RFC approach, but crucial variables were also taken into account. As a result, the precision values rose. As a result, this algorithm was chosen to represent diabetes risk prediction and was employed in the development of the application.

Perveen *et al*.[26] used a data set from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) database to do their research. The study employed the AdaBoost and bagging ensemble techniques using the J48 (C4.5) DT as a base learner and standalone data mining methodology J48 to categorize patients with diabetes mellitus based on diabetes risk indicators. This categorization was done across three separate ordinal adult groups in the CPCSSN. In terms of overall performance, the AdaBoost ensemble method surpassed both bagging and a single J48 DT, according to the findings.

Mujumdar and Vaidehi[27] presented a diabetes prediction model for better diabetes classification that included a few extrinsic factors that caused diabetes, as well as regular components such as glucose, BMI, age, insulin, and so on. The new data set enhanced classification accuracy when compared to the old data set. Multiple ML approaches were used on the data set, and classification was done with a variety of algorithms, with LR yielding the highest accuracy at 96%. The AdaBoost classifier was found to be the most accurate, with a 98.8% accuracy rate. They used two separate data sets to compare the accuracy of ML techniques. When compared to the existing data set, it was clear that the model improved diabetes prediction accuracy and precision.

Mercaldo *et al*.[28] offered a strategy for classifying diabetic patients based on a set of features chosen according to the WHO criteria. Evaluating real-world data using state of the art machine learning algorithms. The model was trained using six alternative classification approaches, with the Hoeffding Tree method scoring 0.770 in precision and 0.775 in recall. They used data from the PIMA Indian community in Phoenix, Arizona, to evaluate the method.

Conclusion

Early detection of diabetes is critical for effective therapy. Many people have no idea whether or not they have it. The full assessment of machine learning approaches for early diabetes prediction and how to apply a variety of supervised and unsupervised machine learning algorithms to the data set to achieve the best accuracy are addressed in this paper. Furthermore, the work will be expanded and refined to create a more precise and general predictive model for diabetes risk prediction at an early stage. Different metrics can be used to assess performance and for accurate diabetic diagnosis.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 1997;20:1183-97.
2. Norris SL, Lau J, Smith SJ, Schmid CH, Engelgau MM. Self-management education for adults with type 2 diabetes: A meta-analysis of the effect on glycemic control. Diabetes Care 2002;25:1159-71.
3. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract 2010;87:4-14.
4. Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, *et al*. Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of Medical Research India Diabetes (ICMRINDIAB) study. Diabetologia 2011;54:3022-7.
5. Ramachandran A, Snehalatha C, Salini J, Vijay V. Use of glimepiride and insulin sensitizers in the treatment of type 2 diabetes-a study in Indians. J Assoc Physicians India 2004;52:459-63.
6. Wagai GA, Romshoo GJ. Adiposity contributes to poor glycemic control in people with diabetes mellitus, a randomized case study, in South Kashmir, India. J Family Med Prim Care 2020;:4623-6.
7. AACE/ACE Position Statement on the Prevention. Diagnosis and treatment of obesity (1998 Revision). Endoc Practice 1998;4:297-330.
8. Birjais R, Mourya AK, Chauhan R, Kaur H. Prediction and diagnosis of future diabetes risk: A machine learning approach. SN Appl Sci 20191:1-8.
9. Sadhu A, Jadli A. Early-stage diabetes risk prediction: A comparative analysis of classification algorithms. Int Adv Res J Sci Eng Technol (IARJSET) 2021;8:193-201.
10. Xue J, Min F, Ma F. Research on diabetes prediction method based on machine learning. J Phys Conf Ser 2020;1684:1-6.
11. Le TM, Vo TM, Pham TN, Dao SV. A novel wrapper–based feature selection for early diabetes prediction enhanced with a metaheuristic. IEEE Access 2020;9:7869-84.
12. Julius AO, Ayokunle AO, Ibrahim FO. Early diabetic risk prediction using machine learning classification techniques. Available from: https://ijisrt.com/early-diabetic-risk-prediction-using-machine-learning-classification-techniques.
13. Shafi S, Ansari GA. Early prediction of diabetes disease & classification of algorithms using machine learning approach. In Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021). Available from: SSRN 3852590 (2021).
14. Khanam JJ, Foo SY. A comparison of machine learning algorithms for diabetes prediction. ICT Express 2021;7:432-9.

15. Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. Procedia Comput Sci 2018;132:1578-85.
16. Agrawal P, Dewangan AK. A brief survey on the techniques used for the diagnosis of diabetes-mellitus. Int Res J Eng Tech IRJET 2015;2:1039-43.
17. Rathore A, Chauhan S, Gujral S. Detecting and predicting diabetes using supervised learning: An approach towards better healthcare for women. Int J Adv Res Comput Sci 2017;8:1192-4.
18. Hassan AS, Malaserene I, Leema AA. Diabetes mellitus prediction using classification techniques. Int J Innov Technol Explor Eng 2020;9:2080-4.
19. Kandhasamy JP, Balamurali S. Performance analysis of classifier models to predict diabetes mellitus. Procedia Comput Sci 2015;47:45-51.
20. Meng XH, Huang YX, Rao DP, Zhang Q, Liu Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. Kaohsiung J Med Sci 2013;29:93-9.
21. Nai-Arun N, Moungmai R. Comparison of classifiers for the risk of diabetes prediction. Procedia Comput Sci 2015;69:132-42.
22. Saravananathan K, Velmurugan T. Analyzing diabetic data using classification algorithms in data mining. Indian J Sci Technol 2016;9:1-6.
23. Kumari VA, Chitra R. Classification of diabetes disease using support vector machine. Int J Eng Res Appl 2013;3:1797-801.
24. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15:104-16.
25. Rawat V, Suryakant S. A classification system for diabetic patients with machine learning techniques. Int J Math Eng Manag Sci 2019;4:729-44.
26. Perveen S, Shahbaz M, Guergachi A, Keshavjee K. Performance analysis of data mining classification techniques to predict diabetes. Procedia Comput Sci 2016;82:115-21.
27. Mujumdar A, Vaidehi V. Diabetes prediction using machine learning algorithms. Procedia Comput Sci 2019;165:292-9.
28. Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Procedia Comput Sci 2017;112:2519-28.