IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Review of Diabetes Disease using Machine Learning Algorithms

[1]UmaDevi P, [2] Padmaja M

[1] Dept of Game Design Technologies, Dr. Ysr Architecture and fine Arts University, Kadapa, Andhra Pradesh, India

[2] Dept of Digital Techniques for Design& Planning, Dr. Ysr Architecture and fine Arts University, Kadapa, Andhra Pradesh, India

Abstract:

Diabetes mellitus (DM) is a chronic disease with a wide range of effects. Diabetes is a disease caused by variables such as age, lack of exercise, sedentary lifestyle, diabetes in the family, high blood pressure, depression and stress, bad diet, and so on. Diabetics are more likely to develop diseases such as heart disease, nerve damage (diabetic neuropathy), eye difficulties (diabetic retinopathy), kidney disease (diabetic nephropathy), stroke, and other complications. Diabetes affects 382 million people globally, according to the International Diabetes Federation. This figure will have climbed to 592 million by 2035. Every day, a great number of people become victims, and many are unaware of their status. It primarily affects people between the ages of between the ages of 25 and 74. Diabetes can cause a plethora of consequences if left untreated and undetected. On the other side, the introduction of machine learning approaches resolves this critical issue.

Keywords: Diabetes mellitus (DM), Diabetic nephropathy , Diabetic retinopathy.

Introduction

Diabetes mellitus is a metabolic condition characterised by unusually high blood sugar levels caused by insulin resistance or a combination of insulin resistance and inadequate insulin production.[1] It is a degenerative metabolic disease that affects all aspects of the patient's life, including physical and emotional well-being, and no treatment method can create dramatic improvements or stop the disease from developing.[2] India has the most diabetics in the world (31.7 million) in 2000, which climbed to 62.4 million in 2011 and is expected to reach 69.9 million by 2025.[3,4] India's high frequency is due to rapid urbanisation and economic development. Indians are

more likely to develop diabetes as a result of their low BMI mixed with diabetes risk factors. High levels of upper-body adiposity, body fat percentage, and insulin resistance.[5] Diabetes symptoms include blurred vision, weight loss, fatigue, increased hunger and thirst, confusion, frequent urination, poor healing, recurrent infections, and difficulties concentrating. "Diabetes is defined as having too much sugar in your blood. High blood sugar problems begin when your body no longer produces enough insulin, a chemical or hormone."[6]
"Sweet urine" is the literal translation. Normal urine contains no sugar. The presence of sugar (or, more precisely, glucose) in the urine is due to an increase in the amount of glucose in the blood, which has spilled over into the urine. Because the body is unable to adequately metabolise glucose, it builds up in the bloodstream. As a result of this, Diabetes is a disease in which the body is unable to appropriately utilise glucose.

Diabetes mellitus (DM)

Diabetes mellitus (DM) is a metabolic disorder caused by a number of factors. It is distinguished by chronic hyperglycemia and changes in carbohydrate, lipid, and protein metabolism caused by insulin shortage, insulin action, or both. Diabetes is a long-term condition. Diabetes, if not controlled properly, can harm neurons and blood vessels in the eyes, kidneys, heart, and lower legs. If blood glucose levels remain high for an extended period of time, complications may arise. Mouth problems include gum disease and tooth decay. Diabetic retinopathy is a visual loss and, in severe situations, blindness disorder. Heart attacks, strokes, and peripheral artery disease (cardiovascular diseases or CVD) (insufficient blood supply to the feet and legs) are all examples of heart and blood vessel ailments. Diabetes nephropathy (kidney disease) is a condition in which the kidneys do not function properly or at all.[7] Type 1 diabetes, type 2 diabetes, and gestational diabetes are the three types of diabetes.

Diabetes type 1 (T1D)

In type 1 diabetes, the body does not create enough insulin. Because body cells cannot absorb glucose from the bloodstream in the absence of insulin, they must rely on alternative sources of energy. Diabetes and its consequences are caused by an excess of glucose in the blood. Insulin- dependent diabetes mellitus (IDDM) is another name for this kind of diabetes. Although it is morecommon in teens and teenagers, it can afflict anyone at any age. It necessitates a complex balancing act of insulin injections (and, in some cases, oral medications), exercise, diet planning, and lifestyle adjustments. Type 1 diabetes symptoms include frequent urination, unusual thirst, unusual hunger, fast weight loss, exhaustion and weakness, nausea, and irritability are a some of the signs of type 1 diabetes.

Diabetes type 2 (T2D)

Type 2 diabetes can manifest as a combination of insulin resistance and secretory dysfunction, with or without insulin resistance. The pancreas generates insulin, but it may be insufficient to maintain normal blood glucose levels, or the cells may be resistant to the insulin produced. The sickness is most common in those over 40, but it is also becoming more common in teenagers and young children. Drowsiness, dry, itchy skin, unwanted weight gain or loss, blurred vision, tingling, numbness, pain in the lower legs, easy fatigue, sluggish healing of wounds

or scratches, and frequent infections (e.g., vaginal infections) are all symptoms of type 1 diabetes. Food, movement, lifestyle control, and, in some cases, oral medications or insulin, are all required.

Neonatal diabetes

Neonatal diabetes is diagnosed in pregnant women who have never had diabetes before but have high blood glucose (sugar) levels during pregnancy. It is a short-term condition that affects 2%- 4% of all pregnant women and normally goes away after the baby is born. Women who have previously experienced gestational diabetes are more likely to develop type 2 diabetes later inlife. This kind of diabetes has no recognized aetiology. The placenta aids in the development of the infant; placental hormones aid in the development of the baby, but they also inhibit the mother's insulin from acting correctly in her body, resulting in insulin resistance. Gestational diabetes develops when a mother's body is unable to produce and use all of the insulin required during pregnancy. The majority of women are completely unaware of any indications or symptoms of gestational diabetes. Two signs are increased thirst and more frequent urination.

Diabetes mellitus is a fatal disease if not treated early; however, early detection can considerably reduce the risk. For early diagnosis, a variety of medical diagnostic methods are already in use. Machine learning algorithms can be used to make early risk forecasts. Recent study on diabetes mellitus risk prediction has yielded promising findings. Machine learning is the study of algorithms that are used to educate machines without the need of people. We can train them to execute a specific job and then use that training to handle comparable tasks without having to explicitly programme them. In medical science, accuracy is always a big issue, and different algorithms may produce varying degrees of accuracy on the same data set. It is critical to determine which algorithm produces the best results while designing a better classifier for better classification. Machine learning is now used in nearly every business. Its application in medical science has the potential to significantly improve healthcare.

Machine learning and classification methods that work well in risk prediction include decision trees, random forests, support vector machines, naive Bayes classifiers, and artificial neuralnetworks. This is achievable due of the algorithms' processing and data handling abilities. Classification accuracy measures can be used to identify the optimal algorithm and determine classification accuracy. This statistic, however, is insufficient to choose the optimal strategy in a proper and efficient manner. Other variables, including as the receiver operating characteristic (ROC) value, F-score, and computation time, should be considered when selecting the optimum conclusion. Classification accuracy, F-score, ROC value, and computation time are all metrics. The findings of this study will benefit future researchers in developing a baseline technique for DM categorization.

Machine Learning Algorithms

Machine learning (ML) is a rapidly evolving field that is being used in a wide range of medical applications. All machine learning models learn from the past and create predictions based on a data set. Diabetes detection will become considerably easier and less expensive as a result of recent developments in machine learning. There are several diabetic data sets available. As a result, machine learning (ML) is necessary for medical diagnoses. The purpose of this study is to predict a patient's risk of getting diabetes. Machine learning algorithms are used. For the study, there are two forms of learning.

1) Supervised Learning
2) Unsupervised Learning.

A supervised learning algorithm's purpose is to predict based on labelled data. The data is labelled in supervised learning. It mimics what a student could learn from a teacher. In contrast, unsupervised learning does not label the data. It's more like self-learning based on prior knowledge. The purpose is to forecast the value of a variable. The data is represented by a set of qualities and attributes. The outcome of directed learning is fixed. Some of the most widely utilised techniques are decision trees (DT), random forests, linear regression, logistic regression, naive Bayes classifiers, k-nearest neighbours (k-NN), support vector machine (SVM), and artificial neural networks (ANN).

Unsupervised learning data consists of values without labels, and the outcome is not predefined The model produces predictions based on self-learning. These models' primary goals are to forecast, categorise, detect, segment, and categorise data. Analysis, recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics are all examples of machine learning applications.

A Review of the Diabetes Prediction on Machine Learning Approaches

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

Sadhu, A., and Jadli[9] conducted experiments on diabetes data from the UCI repository. In total, there were 520 instances and 16 qualities. They aimed to focus their efforts on predicting diabetes early on. Seven classification approaches were used on the validation set of the employed data set: k-NN, logistic regression, SVM, naive Bayes, decision tree, random forests, and multilayer perceptron. According to the results of training several machine learning models, 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 outcomes of training numerous machine learning models.

Xue et al.[10] conducted experiments using a diabetic data set obtained from the UCI repository, which had 520 patients and 17 attributes. They aimed to focus on diabetes detection early on. They used supervised machine learning techniques such as SVM, naive Bayes classifiers, and LightGBM to train on actual data from 520 diabetic and suspected diabetic patients aged 16 to
90. When comparing classification and recognition accuracy, the SVM performs superior. With an accuracy of 93.27%, the naive Bayes classifier is the most extensively used classification algorithm. The highest accuracy rate is 96.54% for SVM. LightGBM is only 88.46% accurate. This reveals that SVM is the best classification algorithm for predicting diabetes.

Le et al.[11] conducted research on early-stage diabetes risk prediction; the data set used in this study was obtained from the UCI repository and included 520 patients and 16 factors. They proposed a machine learning strategy for predicting diabetes onset in patients. It was a novel wrapper-based feature selection strategy that used the grey wolf optimizer (GWO) and adaptive particle swarm optimisation (APSO) to optimize the multilayer perceptron (MLP) and reducethe amount of input attributes required. They also compared the outcomes of this method tothose of classic machine learning methods such as SVM, DT, k-NN, naïve Bayes classifier (NBC), random forest classifier (RFC), and logistic regression (LR). LR attained an accuracy rating of 95%. SVM had a 96% accuracy rate, while k-NN had a 96% accuracy rate.

SVM was 95% accurate, NBC was 93% accurate, DT was 95% accurate, and RFC was 96% accurate. The computational results of the proposed methods show that not only are fewer characteristics required, but also that higher prediction accuracy can be achieved (96% for GWO-MLP and 97% for APSO-MLP). This study has the potential to be used in clinical practice as a tool to help doctors and physicians.

Julius et al.[12] tested a data set obtained from the UCI repository using the Waikato Environment for Knowledge Analysis (Weka) application platform. The data set had 520 samples, each with a set of 17 properties. The purpose of this work was to predict diabetes at an early stage using machine learning classification algorithms based on observable sample features. As classifiers, k-NN, SVM, functional tree (FT), and RFCs were used. The highest accuracy was achieved with k-NN (98%), followed by SVM (94%), FT (93%), and RF (97%).

Diabetes is a dangerous illness, and early detection is always difficult, according to Shafi et al. [13]. This study employed machine learning classification methods to create a model that could handle any problem and might be used to detect diabetes onset at an early stage. The study's authors worked hard to create a framework that could reliably predict the likelihood of diabetes in patients. The three ML approach classification algorithms DT, SVM,

and NBC were tested and evaluated on various measures as part of this study. The PID data set obtained from the UCI repository was used in the study to save time and achieve precise results. According to the experimental data, the NBC technique was adequate, with a 74% accuracy, followed by SVM with a 74% accuracy. SVM came in second with a 63% accuracy while DT came in third with a 72% accuracy. The constructed infrastructure, as well as the ML classifiers used, could be usedto identify or diagnose different diseases in the future. The study, as well as several other ML

as bagging and boosting. RFC performed best in terms of accuracy and ROC score, with an accuracy of 85.558% and a ROC score of 0.912 approaches, might be expanded and enhanced for diabetes research, and the researchers planned to classify other algorithms with missing data.

Khanam et al. [14] investigated diabetes disease prediction. Diabetes is a disease with no known cure, therefore early detection is critical. To predict diabetes, data mining, machine learning (ML), and neural network (NN) technologies were used in this study. They devised a methodfor accurately predicting diabetes. They used data from the PID data collection in the UCI repository. The data collection contained information on 768 patients and their ninecharacteristics. They used seven ML algorithms to predict diabetes on the data set: DT, k-NN, RFC, NBC, AB, LR, and SVM. They preprocessed the data using the Weka programme. They discovered that a model that combines LR and SVM is useful for predicting diabetes. They builta NN model with two hidden layers and varying epochs and discovered that the NN with two hidden layers has an accuracy of 88.6%. ANN received an 88.57%, LR received a 78.85%, NBC received a 78.28%, and RFC received a 77.34%.

Sisodia et al. [15] used a PID data collection from the UCI repository. There were 768 patients and 8 attributes in this data collection. To identify diabetes patients, they used three ML classifications: DT, SVM, and NBC. When compared to the other models, NBC had the highest accuracy (76.30%).In their study, Agarwal et al.[16] also employed the PID data set of 738 patients. The authors used models such as SVM, k-NN, NBC, ID3, C4.5, and CART to assessthe usefulness of this data set in identifying diabetic patients. With an accuracy of 88%, the SVMand LDA algorithms were the most accurate. Rathore et al. [17] used SVM and DTs classification algorithms to predict diabetes mellitus. The data for this inquiry came from the PID data set. PIMA India places a premium on women's health. The SVM has an 82% accuracy. Hassan et al.[18] used classification algorithms such as the DT, k-NN, and SVM to predict diabetes mellitus. With a maximum accuracy of 90.23%, the SVM approach surpassed the DT and KNN algorithms.

On the diabetic data set, Kandhasamy and Balamurali[19] tested the prediction accuracy of J48, k-NN, RFC, and SVM. The author discovered that the J48 approach had a greater accuracy than others, at 73.82%, before preprocessing the data. k-NN and RFC showed improved accuracy after preprocessing.

On the diabetic data set, Meng et al.[20] investigated the J48, LR, and k-NN algorithms. With a classification accuracy of 78.27%, J48 was judged to be the most accurate.

Nai-Arun and Moungmai [21] developed a web application for diabetes prediction based on prediction accuracy. They compared DTs, NNs, LR, NBC, and RFC prediction methods.

Conclusion

Diabetes identification is crucial for optimal treatment. Many people have no idea if they haveit or not. This study addresses the entire review of machine learning approaches for early diabetes prediction, as well as how to use a number of supervised and unsupervised machine learning algorithms to the data set to get the best accuracy. In addition, the work will be developed and modified in order to develop a more precise and general predictive model for diabetes risk prediction at an early stage. Different indicators can be used to evaluate performance and accurately diagnose diabetics.

References:

1) Report of the expert committee on the diagnosis and classification of diabetesmellitus. 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 2010and 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. 2019;1:1–8

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.

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;8:11924.

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 predictdiabetes 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.