Classification of Machine Learning Techniques for Diabetic Diseases Prediction

Sheetal Mahlan[1†] and Sukhvinder Singh Deora[2††],
sheetalmahlan@gmail.com sukhvinder.dcsa@mdurohtak.ac.in
Maharshi Dayanand University, Rohtak, India

Abstract
Diabetes is a condition that can be brought on by a variety of different factors, some of which include, but are not limited to, the following: age, a lack of physical activity, a sedentary lifestyle, a family history of diabetes, high blood pressure, depression and stress, inappropriate eating habits, and so on. Diabetes is a disorder that can be brought on by a number of different factors. A chronic disorder that may lead to a wide range of complications. Diabetes mellitus is synonymous with diabetes. There is a correlation between diabetes and an increased chance of having a variety of various ailments, some of which include, but are not limited to, cardiovascular disease, nerve damage, and eye difficulties. There are a number of illnesses that are connected to kidney dysfunction, including stroke. According to the figures provided by the International Diabetes Federation, there are more than 382 million people all over the world who are afflicted with diabetes. This number will have risen during the years in order to reach 592 million by the year 2035. There are a substantial number of people who become victims on a regular basis, and a significant percentage of those people are uninformed of whether or not they have it. The individuals who are most adversely impacted by it are those who are between the ages of 25 and 74 years old. This paper reviews about various machine learning techniques used to detect diabetes mellitus.

Keywords:
Diabetes mellitus; diabetic neuropathy; diabetic retinopathy;

1. Introduction

According to the description of diabetes mellitus (DM), which can be found in the medical dictionary, diabetes mellitus is "a group of metabolic diseases characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both." There is a connection between persistent hyperglycemia and the metabolic changes that are associated with diabetes mellitus. These changes include glucose, lipid, and protein metabolism [1]. Specific, these organs are more susceptible to the complications that are connected with diabetic complications.

Type 1 diabetes (known as T1DM), type 2 diabetes (known as T2DM), and gestational diabetes are the three forms of diabetes that are defined according to the etiology and clinical presentation of the disease. In the majority of instances, type 1 diabetes mellitus (T1DM) is a consequence of an absolute insulin deficit. This shortfall is a result of the death of β cells in the pancreas, which is mostly caused by an autoimmune process that is mediated by cells. Diabetes mellitus, often known as gestational diabetes, is a disorder that is diagnosed or becomes recognized for the first-time during pregnancy. The intensity of glucose intolerance that is characteristic of this illness may range from mild to severe [2].

This kind of diabetes is the most prevalent, accounting for ninety percent of all diabetic diagnoses. Individuals who are diagnosed with type 2 diabetes often do so after the age of 40; however, this kind of diabetes may also afflict younger individuals and even children. There is a possibility that the afflicted individual may not have symptoms of this sort for a considerable amount of time, and a significant number of patients are identified by accident when they seek treatment for related or unrelated issues. These individuals do not need insulin treatment, in contrast to those who have type 1 diabetes; The use of oral hypoglycemic medications or through the use of diet alone [3]. Individuals diagnosed with type 2 diabetes do not need insulin therapy.

A person's diet, smoking habits, obesity, and lack of physical exercise are all examples of certain risk factors. The term "modifiable" risk factors is occasionally used to refer to behavioral risk factors [4]. This is due to the fact that these risk factors are capable of being adjusted or modified. Diabetes type 2 is one of the noncommunicable diseases (NCDs) that is among the most frequent all over the world, and its

Manuscript received December 5, 2023
Manuscript revised December 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.12.21

prevalence is gradually expanding. This is a fact that is well known. There were more than 460 million individuals who were regarded to be living with diabetes in the year 2019, according to research that was published by the International Diabetes Federation (IDF). According to projections, this number will have climbed to 578 million by the year 2030, and by the year 2045, it is anticipated to have reached 700 million. These figures are supported by estimates. According to estimates provided by the International Diabetes Federation (IDF) [5].

During the course of the past twenty years, a lot of study have been carried out with the objective of forecasting the development of diabetes and its prevalence on a worldwide scale. The research conducted in these studies has used a wide range of data and techniques of analysis [6,7]. The estimates that will be made in the future regarding the burden of diabetes are of the utmost significance for the design of health policy and the evaluation of the expenses that are required to manage diabetes [8,9,10]. Alphabets are used in a wide range of fields, including data modelling, analysis, and visualization [11,12]. Machine learning is a study that focuses on the application of algorithms. The modelling of diabetes has been the subject of a number of research that have been published [13,14,15,16]. Classes 1 and 2 are the lowest, as well as classes 3 and 3 respectively. Class 1 is the lowest class, followed by Class 2, which is the medium low, Class 3, which is the center, and Class 4, which is the highest class. Class 1 is now the lowest class.

The paper is structured in the following manner. In Section 2, a literature review is presented. In the third section, modelling approaches are discussed. Methodology for conducting experiments is covered in Section 4. Part 5 is where the findings are presented. In Section 6, a description of the outcomes that were achieved is also included. Last but not least, Section 7 brings this article to a close and indicates potential topics for further investigation.

In spite of this, it was estimated that about 537 million persons throughout the globe were diagnosed with diabetes in the year 2021. This statistic represents one in ten adults who are diabetic around the world. The International Diabetes Federation (IDF) conducted research that anticipated the number of diabetics in the globe to reach 643 million by the year 2030 and 784 million by the year 2045. As can be seen in Figure 1, the Western Pacific area is now home to the highest number of diabetes patients in the whole globe [5].

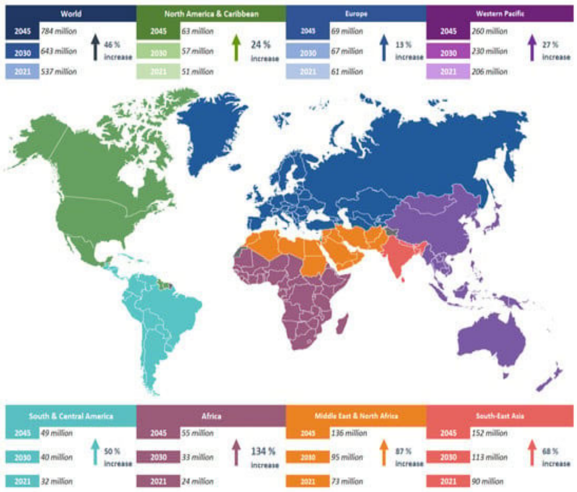


Figure 1. Global diabetes prevalence by region.

1.1 Diabetes and its Types

In the metabolic problem that is often referred to as diabetes mellitus (DM), there are a number of different probable factors that might contribute to the development of the condition. a lack of insulin, or both of these conditions. A kind of ailment that persists for a considerable amount of time is diabetes. In the case that the levels of glucose in the blood continue to be elevated for a lengthy period of time, there is a potential that complications may arise. On the other hand, gum disease and tooth decay are two instances of the many kinds of oral health problems that a person could have. The illness that is known as diabetic retinopathy is one that has the potential to cause a loss of vision and, in more severe cases, blindness among diabetes patients. Circulatory illnesses, sometimes referred to as cardiovascular diseases or CVD, include conditions such as heart attacks, strokes, and peripheral artery disease. These are all instances of ailments that affect the heart and blood arteries.

Because of these problems, the distribution of blood to the legs and feet is insufficient. Chronic kidney disease, also known as diabetic nephropathy, is marked by the kidneys not working properly or at all. This condition is also often referred to as kidney disease.[7]: Diabetes manifests itself in three distinct forms: type 1 diabetes, type 2 diabetes, and gestational diabetes. These three kinds of diabetes are all considered to be subtypes of diabetes.

Type 1 diabetes (T1D)

When a person has type 1 diabetes, their body struggles to create enough insulin. Because insulin is required for the cells of the body to be able to absorb glucose from the circulation, the cells are forced to depend on alternative sources of energy. Diabetes and the difficulties that come along with it are caused by an excess of glucose in the blood. Another name for this kind of diabetes is insulin-dependent diabetes mellitus (IDDM), which is an abbreviation. People of any age may be affected by it, however adolescents and teens are more likely to be affected than other age groups. Injections of insulin (and, in some cases, oral medications), physical activity, careful planning of meals, and adjustments to one's lifestyle are all necessary components of this method of managing diabetes.

Type 2 diabetes (T2D)

Over the age of forty is the age group that is most likely to be affected by the disease; nevertheless, adolescents and young children are also showing signs of increasing prevalence. Type 2 diabetes, you will need to regulate your lifestyle, eat well, and exercise regularly. In certain cases, you may also need to take insulin or oral medications.[6]

1.2 Machine Learning Algorithms

The area of machine learning (ML) is one that is presently seeing significant advancement and is being used in the medical business for a wide range of reasons at the present time.[45] The models that are used for machine learning all learn from their prior experiences and form conclusions depending on the data that they have to work with. As a consequence of recent advancements in machine learning, the process of diagnosing diabetes will become far less challenging and more cost-effective. People who have diabetes have access to a large number of data sets relevant to the condition. Because of this, machine learning is an essential tool for the diagnosis of medical conditions. This study has a number of goals, one of which is to produce a forecast regarding the chance of a patient developing diabetes. The application of algorithms is what is meant by the term "machine learning." There are two aspects of learning that are relevant to the study that is being conducted.

(1) Observation and Instruction
(2) The process of learning without supervision.

The purpose of an algorithm for supervised learning is to make predictions based on data that has been labelled. Labelling the data is an essential part of supervised learning. The simulation is a representation of what a learner may learn from a teacher.

It's more like learning on your own based on the experiences you've had in the past. The objective is to develop a value prediction for a variable. To represent the data, a collection of characteristics and characteristics are employed. There is no room for doubt about the consequence of directed learning. The model is able to generate predictions since it is based on self-learning. The most important objectives of these models are to achieve forecasting, classification, detection, segmentation, and categorization of data. The following are some examples of applications that may be made use of machine learning: analysis, recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics.

2. Diabetes Prediction Literature Review Using Machine Learning

The accuracy of naive Bayes was found to be 77%, while the accuracy of logistic regression was found to be 79%. Gradient boosting was found to have an accuracy of 86%, while naive Bayes had an accuracy of 77%.

Using a diabetic data set that was obtained from the UCI repository, Sadhu, A. and Jadli, A.[9] conducted an experiment. Seven different classification methods were used to the validation set of the data set that was used. These methods were k-nearest neighbours, logistic regression, support vector machines, naive Bayes, decision trees, random forests, and multilayer perceptron.

The use of supervised machine learning methods including support vector machines (SVM), naive Bayes classifiers, and Light GBM, they trained on the real data of 520 diabetes patients and probable diabetic patients ranging in age from 16 to 90. According to the results of a comparison between the accuracy of classification and recognition, the performance of the SVM is the most striking.

With an accuracy rate of 93.27%, the naive Bayes classifier is the classification technique that is used the most often now. Based on its accuracy rating of 96.54%, SVM is the most accurate. In terms of accuracy, Light GBM is only 88.46% accurate. Based on this evidence, support vector machines (SVM) are the most effective classification technique for diabetes prediction.

The repository at the University of California, Irvine (UCI) was the source of the data set that was used in this investigation. The results achieved using this technique were also compared to the results obtained using a number of classic machine learning algorithms, such as support vector machines (SVM), deep neural networks (DT), k-nearest neighbours (k-NN), naïve Bayes classifier (NBC), random forest classifier (RFC), and logistic regression methods (LR). A better level of prediction accuracy (96% for GWO–MLP and 97% for APSO–MLP are the respective figures). The findings of this study have the potential to be used in clinical settings and to serve as a resource for medical professionals and practitioners.

The PID data set that was accessible via the UCI repository was used by Sisodia et al. [15]. A total of 768 patients and eight characteristics were included in this data collection. In order to identify diabetes individuals, they used three different machine learning classifications: DT, SVM, and NBC. When compared to the other models, NBC had the greatest accuracy, which was calculated to be 76.30%.

The DT, k-NN, and SVM are examples of classification methods that were used by Hassan et al. [18] in order to make a prediction about diabetes mellitus. With a maximum accuracy of 90.23 percent, the SVM algorithm fared better than both the DT and KNN approaches.

Author found out that the J48 approach had a greater accuracy than other methods, with a score of 73.82%, before they carried out any preprocessing on the data. Following the completion of preprocessing, both k-NN and RFC showed enhanced accuracy.
J48, LR, and k-NN algorithms were evaluated by Meng et al. [20] using the diabetic data set as their subject. The classification accuracy of J48 was judged to be the highest, coming in at 78.27% after being evaluated.

Nai-Arun and Moungmai [21] developed an online application for diabetes prediction that is dependent on the accuracy of the forecast. A number of different prediction approaches, including bagging and boosting, were compared by them.

In order to achieve their goal of making a forecast about diabetes, Kavakiotis et al. [22,23] used the NBC, RFC, k-NN, SVM, DT, and LR methodologies. There was a tenfold cross-validation approach that was used in the application of the algorithms. According to the findings of the research, the SVM performed the best in terms of accuracy among all the methods, measuring 84%.

Table 1 Features comparison of planned study to state-of-the-art studies.

Study	Diabetes classification	Diabetes prediction	Real-time healthcare data analysis	Performance measures
Proposed	✓	✓	✓	Accuracy, Precision, Recall, RMSE, r
[8]	✓	✕	✓	Accuracy, standard deviation
[34]	✓	✕	✓	Accuracy
[26]	✕	✓	✕	Accuracy, correlation coefficient
[4]	✕	✕	✓	NA

3. Diabetes Classification for Healthcare

The process of diagnosing a patient's health status is not only necessary but also very important for medical experts. The classification of a diabetes type is one of the most complicated occurrences for medical experts, and it involves a number of different tests. On the other hand, doing a diagnostic that takes into account a number of different aspects might sometimes result in erroneous findings. Consequently, the interpretation and categorization of diabetes are tasks that are very difficult to do. The healthcare industry has benefited significantly from recent technological breakthroughs, especially those that apply to machine learning approaches. These improvements have demonstrated to be highly useful. The classification of diabetes has been the subject of a significant deal of study, and the research literature has been published with a great deal of different approaches.[46] An examination of photoplethysmograms was used to develop a logistic regression model that was proposed by Qawqzeh and colleagues for the categorization of diabetes. For training purposes, they utilized the data from 459 patients, and for testing and validating the model, they used 128 data points. They were able to attain an accuracy of 92% using their suggested approach, which accurately identified 552 individuals as not having diabetes.

4. What the classifiers' theoretical ideas are

For the sake of your convenience, the several classifiers that are used have been segregated into their respective sub-sections respectively.

The Regression of Logistic

This is done by the use of a statistical model that is founded on logical function, which allows for the development of a binary-dependent variable. According to Diwani and Sam (2014), probabilities are used in order to produce an estimate of the relationship that exists between the variables that are dependent and those that are associated with the independent variables.

When using this strategy, the dependent variable is a category variable. According to Kaur and Chhabra (2014), the mathematical expression for it is shown as follows:

$$h_\theta(x) = P(Y = 1 \mid X;\ \text{thet}\ \alpha) \quad (1)$$

The likelihood that Y equals 1 given X, which is denoted by the symbol "theta"

$$\text{P}(\text{Y} = 1 \mid \text{X};\ \text{theta}) + \text{P}(\text{Y} = 0 \mid \text{X};\ \text{theta}) = 1 \quad (2)$$

This is the XGBoost.

The installation of gradient enhanced DTs that are brought into existence in a sequential fashion is what it is. Its weights are an essential factor to consider. For the purpose of obtaining the findings, the DTs are given a specific weight that is applied to each individual variable (Butt et al., 2021). Each individual

DT's prediction scores are determined using the following formula:

$$\hat{y}_i = \sum_{k=1}^{K} f_k \in F \quad (3)$$

In the study by Patil et al. (2019), the sign k is used to indicate the number of trees, the symbol f is used to describe the functional space, and the symbol F is used to illustrate the potential set that is accessible.

Gradient Boosting

According to the findings of a study that was conducted in the year 2020 by Sehly and Mezher, the majority of students who are failing to learn are included into a predictive model, which often takes the shape of DTs.

It is primarily used in situations in which we want to reduce the bias mistake. Based on Posonia et al.'s 2020 research, a method known as gradient-descent is used for the purpose of obtaining the coefficient values.

The function that is used for loss is (y1 - y1′)2. The actual value is denoted by y1, whereas the value that is predicted by this model is denoted by y1′. According to Ke et al. (2017), the true goal is represented by Gn(X), which is therefore substituted for y1′. Specifically, it may be stated numerically as follows:

$$\begin{aligned} G_{n+1}(X) &= G_n(X) + \gamma_n H1(x, e_n) \\ \text{L1} &= (y1 - y1')^2 \\ \text{L1} &= (Y - G_n(x))^2 \end{aligned} \quad (4)$$

Decision Trees

Depending on the situation. According to Chen and Guestrin (2016), it is used to indicate if the processes in question belong to the classification or regression categories. There are a few different types of DTs that are now accessible, and they include ID3, ID 4.5, CART, and CHAID.

Following is a list of the measurements that are utilized on DT: In Khanam and Foo's 2021, entropy, the Gini index, and standard deviation were all considered. The following is the mathematical calculation that is used to determine it (Ambigavathi and Sridharan, 2018):

$$\begin{aligned} \text{Entropy } &= -\sum_{i1=1}^{n1} p_{i1}^{*} \log(p_{i1}) \\ \text{Gini Index } &= 1 - \sum_{i=1}^{n1} p_{i1}^{2} \end{aligned} \quad (5)$$

Extra Trees

Additional trees, often known as ETs, are frequently referred to as "extremely randomized trees classifiers." According to Chen et al. (2017) RF in that the construction of DTs is different. The entropy is determined by the following formula:

$\text{Entropy (S1)} = \sum_{\text{i1}=1}^{\text{c1}} -\text{p}_{\text{i1}}\log_2 \text{ p}_{\text{i1}}$

$$\text{Entropy (S1)} = \sum_{\text{i1}=1}^{\text{c1}} -\text{p}_{\text{i1}}\log_2 \text{ p}_{\text{i1}} \quad (6)$$

where the number of distinct class labels is denoted by the symbol c1, and the percentage of rows that include an output label is denoted by the symbol pi1 (Sisodia and Sisodia, 2018).

$$\text{Gain } (S1, A) = \text{Entropy } (S1) - \sum_{v\varepsilon \text{ Value } (A)} \frac{|S1v|}{|S1|} \text{Entropy } (S1_V) \quad (7)$$

Light Gradient Booster

It is referred to as a "GB framework" that is created on the DT algorithm, as stated by Ahmed and Arya (2021). The most important functions that it performs are those of classification and ranking. The tree is chopped down one leaf at a time using the best-fit method. This method is used to cut everything down individually. According to the conclusions of the research that was done by Zhu and colleagues in the year 2020, Several various approaches might be used to communicate this idea, including the following:

TABLE 1. Percentage accuracy.

Date set	Logistic regression	XGB classifier	Gradient Boosting classifier	Decision tree	Extra trees Classifier	Random forest	LGBM
PIMA indian Dataset	75.20%	83.30 %	94.10 %	94.40 %	94.60%	94.80 %	95.20 %

The LGBM algorithm, which has a rate of accuracy of 95.2%, is capable of achieving the maximum degree of precision that is possible. Comparison with the previous methods is shown in table 2.

Table 2 Comparison with the previous works

Previous work	Method	Accuracy(%)
Zolfaghari [33]	Ensemble of SVM and NN	88.04
Sneha and Gangil [34]	Feature Selection and SVM	77.37
Edeh, Khalaf, Tavera, Tayeb, Ghouali, Abdulsahib, Richard – Nnabu and [35]	SVM	83.1
Massaro, Maritati, Giannone, Convertini and Galiano [36]	LSTM-AR	89
Dadgar and Kaardaan [37]	UTA-NN and GA	87.46
Zou, QU, Luo, Yin, JU and Tang [38]	Mrmr-rf	77.21
Ashiquzzaman, Tushar, Islam, Shon, Im, Park, Lim and Kim [39]	Deep MLP	88.41
Kannadasan, Edla and Kuppili [40]	Stacked Autoencoders-DNN	86.26
Rahman, Islam, Mukti and Saha [41]	Conv-LSTM	91.38
Alex, Nayahi, Shine and Gopirekha [42]	DCNN/SMOTE/Outlier Detection	86.29

5. Conclusion

From the above study, it can be concluded that the LGBM algorithm is capable of achieving the maximum degree of precision in comparison of various machine learning algorithms. The accuracy of the predicted diabetes mellitus disease can be improved by augmenting the dataset with other sophisticated techniques such as transformer-based learning. Attributes can also be used in various combinations for identification purposes. The classifiers can be further fine-tuned to predict the disease more accurately. The probability of occurrence of diabetes mellitus disease can also be calculated. In order to circumvent the challenges of employing linear functions in high-dimensional feature space, SVM and support vector regression may be employed to change the optimization issue into dual convex quadratic approaches. This further improves the accuracy percentage and provides a deeper model for predicting diabetes mellitus disease in affected people. When data points cannot be partitioned along a linear axis, SVM uses a high-dimensional feature space to categories them. After determining where the two groups part ways, the information is transformed such that the boundary may be shown visually as a hyper plane.

References

[1] American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2006, 29, S43. [Google Scholar] [CrossRef]

[2] Jamison, D.T.; Breman, J.G.; Measham, A.R.; Alleyne, G.; Claeson, M.; Evans, D.B.; Jha, P.; Mills, A.; Musgrove, P. Disease Control Priorities in Developing Countries; World Bank Publications: Washington, DC, USA, 2006. [Google Scholar]

[3] World Health Organization (WHO). Diabetes Country Profiles 2016. 2016. Available online: https://cdn.who.int/media/docs/default-source/ncds/ncd-surveillance/diabetes_profiles_explanatory_notes.pdf?sfvrsn=f2a2083c_5&download=true (accessed on 1 July 2022).

[4] Rewers, M.; Hamman, R.F. Risk factors for non-insulin-dependent diabetes. Diabetes Am. 1995, 2, 179–220. [Google Scholar]

[5] International Diabetes Federation (IDF). IDF Diabetes Atlas, 9th ed.; International Diabetes Federation: Brussels, Belgium, 2019. [Google Scholar]

[6] Guariguata, L.; Whiting, D.R.; Hambleton, I.; Beagley, J.; Linnenkamp, U.; Shaw, J.E. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res.

Clin. Pract. 2014, 103, 137–149. [Google Scholar] [CrossRef] [PubMed]

[7] NCD Risk Factor Collaboration (NCD-RisC); Walton, J. Worldwide trends in diabetes since 1980: A pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016, 387, 1513–1530. [Google Scholar] [CrossRef] [PubMed][Green Version]

[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. [Google Scholar]

[9] Sadhu A, Jadli A. Early-stage diabetes risk prediction:A comparative analysis of classification algorithms. IntAdv Res J SciEngTechnol (IARJSET) 2021;8:193–201. [Google Scholar]

[10] Xue J, Min F, Ma F. Research on diabetes prediction method based on machine learning. J PhysConf Ser. 2020;1684:1–6. [Google Scholar]

[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. [Google Scholar]

[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 . [Google Scholar]

[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) [Google Scholar]

[14] Khanam JJ, Foo SY. A comparison of machine learning algorithms for diabetes prediction. ICT Express. 2021;7:432–9. [Google Scholar]

[15] Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. Procedia Comput Sci. 2018;132:1578–85. [Google Scholar]

[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. [Google Scholar]

[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. [Google Scholar]

[18] Hassan AS, Malaserene I, Leema AA. Diabetes mellitus prediction using classification techniques. Int J InnovTechnolExplor Eng. 2020;9:2080–4. [Google Scholar]

[19] Kandhasamy JP, Balamurali S. Performance analysis of classifier models to predict diabetes mellitus. Procedia Comput Sci. 2015;47:45–51. [Google Scholar]

[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. [PubMed] [Google Scholar]

[21] Nai-Arun N, Moungmai R. Comparison of classifiers for the risk of diabetes prediction. Procedia Comput Sci. 2015;69:132–42. [Google Scholar]

[22] Saravananathan K, Velmurugan T. Analyzing diabetic data using classification algorithms in data mining. Indian J Sci Technol. 2016;9:1–6. [Google Scholar]

[23] 23. Kumari VA, Chitra R. Classification of diabetes disease using support vector machine. Int J Eng Res Appl. 2013;3:1797–801. [Google Scholar]

[24] 24. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. ComputStructBiotechnol J. 2017;15:104–16. [PMC free article] [PubMed] [Google Scholar]

[25] Rawat V, Suryakant S. A classification system for diabetic patients with machine learning techniques. Int J Math EngManag Sci. 2019;4:729–44. [Google Scholar]

[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. [Google Scholar]

[27] Mujumdar A, Vaidehi V. Diabetes prediction using machine learning algorithms. Procedia Comput Sci. 2019;165:292–9. [Google Scholar]

[28] Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Procedia Comput Sci. 2017;112:2519–28. [Google Scholar]

[29] S. Wadhwa and K. Babber, "Artificial intelligence in health care: predictive analysis on diabetes using machine learning algorithms," in Proceeding of the International Conference on Computational Science and Its Applications, pp. 354–366, Springer, Cagliari, Italy, July 2020.

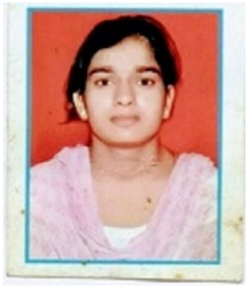
[30] S. Majumder, Y. Elloumi, M. Akil, R. Kachouri, and N. Kehtarnavaz, "A deep learning-based smartphone app for real-time detection of five stages of diabetic retinopathy," in Proceedings of the Real-Time Image Processing and Deep Learning 2020, vol. 11401, p. 1140106, International Society for Optics and Photonics, April 2020.

[31] A. Hussain and S. Naaz, "Prediction of diabetes mellitus: comparative study of various machine learning models," in Proceeding of the International Conference on Innovative Computing and Communications, pp. 103–115, Springer, Delhi, India, January 2021.

[32] G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Calibration of minimally invasive continuous glucose monitoring sensors: state-of-the-art and current perspectives," Biosensors, vol. 8, no. 1, 2018.

[33] Zolfaghari R. Diagnosis of diabetes in female population of pima indian heritage with ensemble of bp neural network and svm. Int. J. Comput. Eng. Manag/ 2012;15:2230–7893. [Google Scholar]

[34] Sneha N., Gangil T. Analysis of diabetes mellitus for early prediction using optimal features selection. J. Big Data. 2019;6:13. doi: 10.1186/s40537-019-0175-6. [CrossRef] [Google Scholar]
[35] Edeh M.O., Khalaf O.I., Tavera C.A., Tayeb S., Ghouali S., Abdulsahib G.M., Richard-Nnabu N.E., Louni A. A Classification Algorithm-Based Hybrid Diabetes Prediction Model. Front. Public Health. 2022;10:829519. doi: 10.3389/fpubh.2022.829519. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
[36] Massaro A., Maritati V., Giannone D., Convertini D., Galiano A. LSTM DSS Automatism and Dataset Optimization for Diabetes Prediction. Appl. Sci. 2019;9:3532. doi: 10.3390/app9173532. [CrossRef] [Google Scholar]
[37] Dadgar S.M.H., Kaardaan M. A Hybrid Method of Feature Selection and Neural Network with Genetic Algorithm to Predict Diabetes. Int. J. Mechatron. Electr. Comput. Technol. (IJMEC) 2017;7:3397–3404. [Google Scholar]
[38] Zou Q., Qu K., Luo Y., Yin D., Ju Y., Tang H. Predicting Diabetes Mellitus With Machine Learning Techniques. Front. Genet. 2018;9:515. doi: 10.3389/fgene.2018.00515. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
[39] Ashiquzzaman A., Tushar A.K., Islam M., Shon D., Im K., Park J.-H., Lim D.-S., Kim J. IT Convergence and Security 2017. Springer; Singapore: 2018. Reduction of overfitting in diabetes prediction using deep learning neural network; pp. 35–43. [Google Scholar]
[40] Kannadasan K., Edla D.R., Kuppili V. Type 2 diabetes data classification using stacked autoencoders in deep neural networks. Clin. Epidemiol. Glob. Health. 2019;7:530–535. doi: 10.1016/j.cegh.2018.12.004. [CrossRef] [Google Scholar]
[41] Rahman M., Islam D., Mukti R.J., Saha I. A deep learning approach based on convolutional LSTM for detecting diabetes. Comput. Biol. Chem. 2020;88:107329. doi: 10.1016/j.compbiolchem.2020.107329. [PubMed] [CrossRef] [Google Scholar]
[42] Alex S.A., Nayahi J., Shine H., Gopirekha V. Deep convolutional neural network for diabetes mellitus prediction. Neural Comput. Appl. 2022;34:1319–1327. doi: 10.1007/s00521-021-06431-7. [CrossRef] [Google Scholar]
[43] Kalagotla S.K., Gangashetty S.V., Giridhar K. A novel stacking technique for prediction of diabetes. Comput. Biol. Med. 2021;135:104554. doi: 10.1016/j.compbiomed.2021.104554. [PubMed] [CrossRef] [Google Scholar]
[44] Jakka A., Vakula Rani J. Performance evaluation of machine learning models for diabetes prediction. Int. J. Innov. Technol. Explor. Eng. (IJITEE) 2019;8:1976–1980. [Google Scholar]
[45] Sheetal, Dr Sukhvinder Singh Deora, "DIABETIC DISEASES PREDICTION USING MACHINE LEARNING TECHNIQUES: A REVIEW" in Proceedings of the National Conference on Computational Intelligence and Data Science (NCCIDS-23), March, 2023, MDU, Rohtak, pp. 195-199.
[46] Sukhvinder Singh Deora, Mandeep Kaur, “Image Processing and Computer Vision: Relevance and Applications in the Modern World” in Nova Science Publishers, 2023, The Impact of Thrust Technologies on Image Processing, https://doi.org/10.52305/ATJL4552 Volume 1 Pages 233-252



Sheetal Mahlan received the BTech and MTech in Computer Science & Engineering from Maharshi Dayanand University in 2015 and 2017 respectively. She is Pursuing her PhD from Maharshi Dayanand University, Rohtak, India. Her interest areas include Machine Learning, Theory of Computation, Data Analytics, and Database design. She has written research articles involving latest machine learning techniques for diseases prediction.

Sukhvinder Singh Deora received the MSc & M.C.A. from Kurukshetra University in 2000 and 2002. He did his M.Phil. in Computer Science and completed PhD in 2015. He is currently working as Assistant Professor in Department of Computer Sciences, Maharshi Dayanand University, Rohtak, India. He is a reviewer to many prestigious International Journals and member of editorial board of some Journals. To his credit are many prominentpapers in the area of data security, big data analytics and issues related to it. He has also edited Proceedings of National Level Seminars/Conferences. With an exposure of 1.5 years in IT industry his thrust areas include Testing, Java technologies and Database design issues. His interest areas include Big Data Analytics, Network Security, Theoretical Computer Sciences, Data Structures, S/W Testing and Database Designing. He is also a member of ACM, Computer Society of India (CSI) and Indian Society of Information Theory and Applications (ISITA).