An evaluation framework for diabetes prediction techniques using machine learning

Aya Ahmed Hashim[1*] and *Ayad* Hameed Mousa[2]

[1] University of Kerbala, Kerbala, Iraq
[2] University of Kerbala, Kerbala, Iraq

Abstract. Diabetes affects a large segment of society and does not discriminate based on age. Children, young people, or the elderly may be affected by it. By detecting the disease early, clinicians can help patients recover or at least control it. Models based on machine learning algorithms have been proposed by researchers in the field of artificial intelligence to predict disease and determine its type. The purpose of this study was to propose a framework for evaluating studies related to diabetes detection and identification. To develop the proposed model, a systematic review of studies related to the topic was conducted. After proposing and evaluating the framework, 54 relevant studies were evaluated and results inspired by it were drawn.

Introduction

Diabetes Mellitus (DM) means a high level of sugar in the blood, as the body becomes unable to deal with the amount of sugar in the bloodstream. It is a serious and chronic disease that affects different age groups. Early detection of diabetes helps control it and avoid exposure to other health problems such as heart and blood vessel diseases. Foot ulcers, blurry eyes, etc [1]

The pancreas produces the hormone insulin, and the work of insulin is to help the glucose present in the bloodstream to move into the body's cells. The development and occurrence of diabetes and the accumulation of glucose in the blood are caused by one of two conditions: either the pancreas cannot produce a sufficient amount of insulin (insulin deficiency) or the body is unable to use the insulin it has produced (insulin resistance)[2]. Diabetes has several types, the main types of which are gestational diabetes, type 1 diabetes T1DM, and type 2 diabetes T2DM
the body produces an insufficient amount of insulin and is also called insulin disease that affects adults.

* Corresponding author: aya.a.hashim@s.uokerbala.edu.iq

T1DM is also known as juvenile diabetes. In this type, the insulin-secreting cells are damaged and insulin production in the body stops.

T2DM is milder than the first type and affects the elderly. This type occurs when gestational diabetes occurs in the second half of pregnancy, when the body is either unable to produce a sufficient amount of insulin or does not use insulin, which leads to high blood sugar and the occurrence of gestational diabetes[3].

Recently, applications of artificial intelligence have developed rapidly in many fields, including the health sector. Technical improvements have led to the emergence of many inventions in the field of artificial intelligence[4].

Machine learning and deep learning are parts of artificial intelligence [3]. Recently, systems based on machine learning have dominated the medical field in all its aspects and various diseases in order to classify and diagnose these diseases [5]. Early diagnosis of diseases is very important, and artificial intelligence techniques provide a reference in obtaining the initial scoop about diseases and thus help in controlling or treating them and reducing risks[6, 7].

Method

The systematic review carried out depending on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines[8]. We prepared this systematic review by searching various online public Libraries, such as IEEEXplore, Elsevier, ACM, Springer and others. Elsevier is a global information analytics and publishing company specializing in scientific, technical, and medical content that helps institutions and professionals progress science, advance healthcare and improve the performance. IEEE Xplore digital library is a research database for discovery and access to journal articles, conference proceedings, technical standards, and related materials on computer science, electrical engineering and electronics, and allied fields. Predictive models for diabetes studies were covering the period June 2017 to February 2023 were searched using the following search string: ((“diabetes” OR “diabetes mellitus”) AND (“prediction model” OR “predictive model” OR “predicting” OR “detection” OR “machine learning algorithm” OR “algorithm”)) All selected studies are in English and are not review. The search was conducted between November 2023 and January 2024, and was depends on titles and abstracts.

 https://doi.org/10.1051/bioconf/20249700125

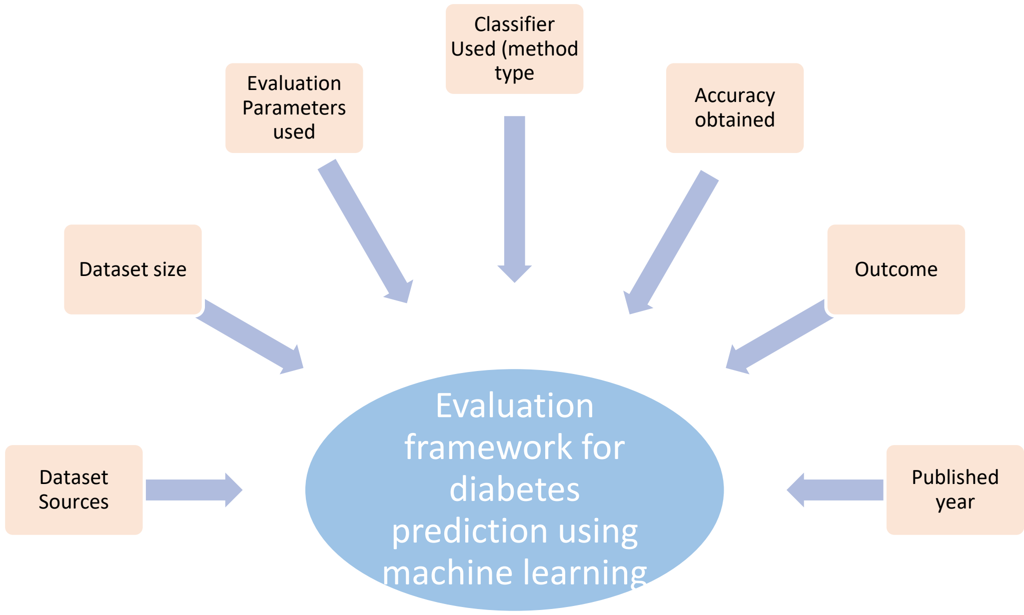

Fig .1. The evaluation framework for diabetes prediction.

DATA COLLECTION PROCESS

The reviewer (AAH) independently screened the titles and abstracts of all articles identified by the search string to exclude articles not related to prediction models. Full texts were retrieved for all records that passed initial screening or if eligibility was indeterminate. Full text articles were independently screened by 2 reviewers (AAH, AHM) against eligibility criteria. In articles that presented more than one model, the model recommended by the authors was chosen. We proposed and evaluate a framework for diabetes prediction using machine learning techniques (Figure 1**)**, the data elements extracted for this review include [Dataset Sources, Dataset size, Evaluation Parameters used, Classifier Used (method type), Accuracy obtained, Outcome and Published year]. We reported our systematic review according to the PRISMA guidelines [9], excluding items which were not needed by our study.

https://doi.org/10.1051/bioconf/20249700125

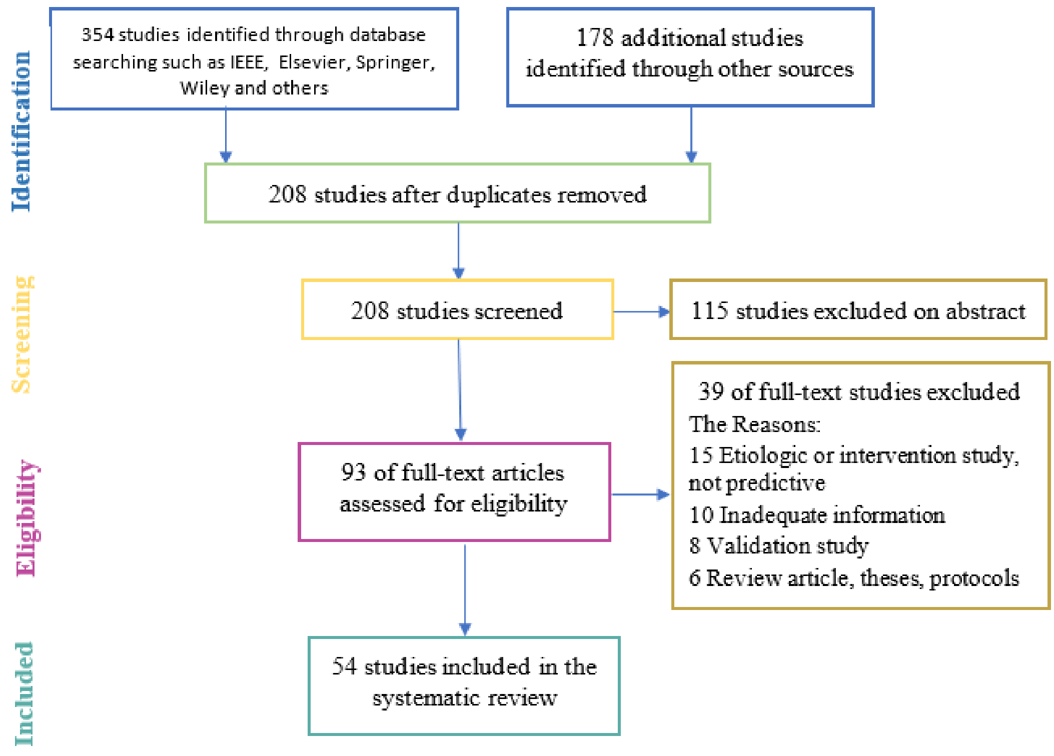

Fig .2. Flow diagram of selected studies.

In addition of, Figure 3 Analyse the distribution of selected research by publication year

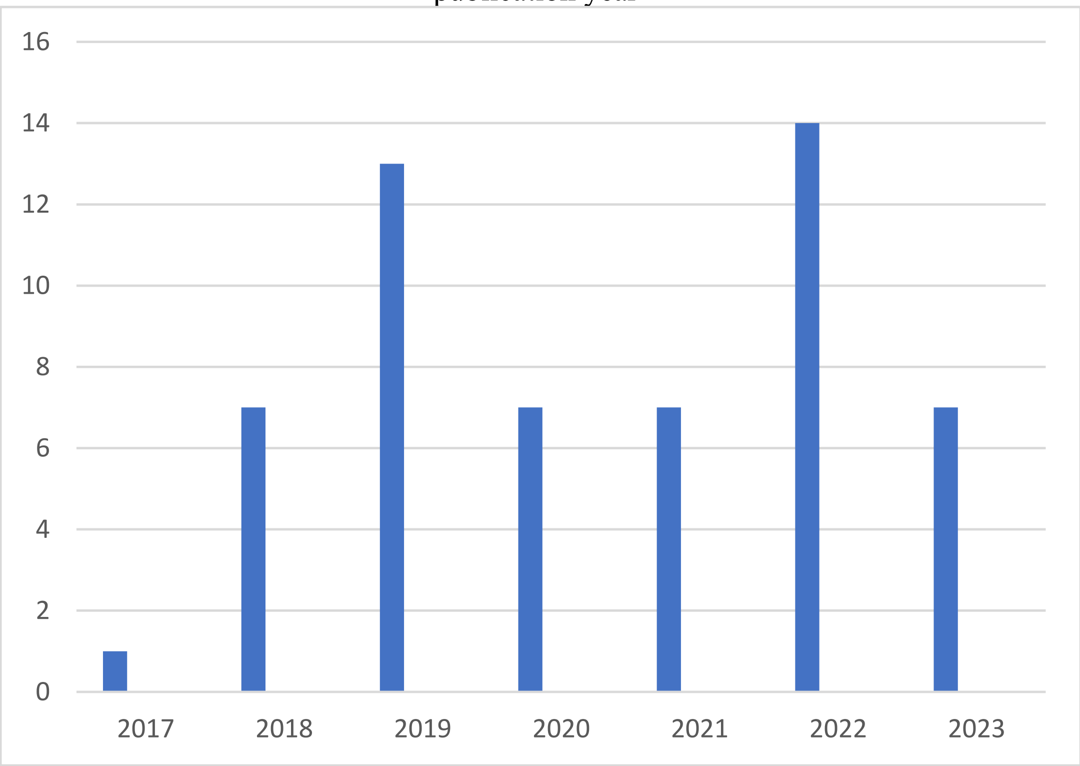

Fig .3. the distribution of selected research by publication year

DATA ITEMS

1 Dataset Sources:

The source of input data is a vital factor for machine learning and deep learning models. Most of these studies used one dataset, and some used more, inclusive public and private datasets, to validate the generalization of machine learning and deep learning models. Thus, this category summarizes information about the datasets used.

2 Dataset size:

This field indicates the number of individuals with and without diabetes present in the indicated dataset.

3 Evaluation Parameters used:

Usually in data sets there are many missing values, outliers, and unbalanced distribution of classes, so the accuracy is not enough to evaluate the performance of the prediction and classification model. For this reason, it is important to take different statistical metrics such as such as F1-score, precision, recall, and Receiver Operating Characteristic Curve (ROC_AUC) into account to measure the performance.

4 Classifier Used (method type):

In this field, we show what model is proposed by the researcher that provided the highest accuracy and performance when compared to a group of different models. Researchers have presented many models in machine learning such as decision tree (DT), Support vector machine (SVM), Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN), Gradient boosting, XGBoost, AdaBoost, LGBM and various ensemble techniques. There are some researchers who have used deep learning to predict diabetes, and among these models are the deep neural network, back propagation and Artificial Neural Network.

5 Accuracy obtained:

Accuracy is a common and important evaluation metric and is the ratio of the total correctly predicted observation to the total number of observations so it is used to measure the performance of any prediction model

6 Outcome:

In this part, we will explain what the purpose is or what the results extracted from the research indicate, whether it is to predict diabetes, predict type 2 diabetes, predict gestational diabetes, or classify these types.

7 Published year:

The search was limited to publications from June 2017 to August 2023 to ensure data were up to date within six years from this study.

RESULTS

In this systematic review, as shown in the (figure 2**)**, by searching for studies, we obtained 532 studies through screening the title, and after removing duplicates, we obtained 208 studies were screened by abstract, and we excluded 115 studies. After this stage, 93 full-text articles assessed for eligibility, and we excluded 39 studies. In total, 54 studies were eligible for review.

Through the studies included in this research, we can see that most of them focused generally on classifying the sample into those with diabetes or those without diabetes.

While other research took it upon itself to determine the type of diabetes, some of this research included identifying type 1 diabetes only, others focused on identifying type 2 diabetes, and another part of the research focused on identifying gestational diabetes.

Conclusion

In this article, we have included a wide range of articles related to diabetes prediction, in line with the above situation, this study concludes that the comprehensive model to cover all diabetes type is missing and investigation of diabetes types identification is significant

Table 1 Summary of selected articles on diabetes prediction using machine learning techniques.

no	Author	Dataset sources	Data size	Evaluation Parameters used	Classifier Used (method type)	Accuracy obtained	outcome	published Year
[9]	Haya Majid Qureshi	Collected from a local hospital.	2000	Sensitivity Specificity	*PCA-SVM*	98%	To classification diabetes type	June 2017
[10]	N. Yuvaraj	data collected by National Institute of Diabetes	75,664	precision, recall, F-Measure	Random forest	94%	To predict diabetes	2018
[11]	Fikirte Girma Woldemichael	Pima Indians Diabetes Dataset	768	Sensitivity, Specificity	Back propagation algorithm	83.11%	To predict diabetes	2018
[12]	Suyash Srivastava	Pima Indians Diabetes Dataset	768	(RMSE)and ROC	Artificial Neural Network	92%	To predict diabetes	2018
[13]	Quan Zou	from hospital physical examination data in Luzhou, China.and Pima Indians Diabetes Dataset	68994, 768	sensitivity (SN), specificity (SP), and Matthews correlation coefficient (MCC)	RF	80.84% 77.21%	To predict diabetes	6 November 2018
[14]	Ayman Mir	Pima Indians Diabetes Dataset	768	TP-rate, FP-rate, precision, recall, F-Measure.	Support Vector Machine	79.13%	To predict diabetes	2018
[15]	Deepti Sisodiaa	Pima Indians Diabetes Dataset	768	Precision, Recall, F-Measure, ROC	Naive Bayes	76.30 %	To predict diabetes	2018
[16]	Muhammad Azeem Sarwar,	Pima Indians Diabetes Dataset	768	TP, FN, FP, TN	SVM and KNN	77%	To predict diabetes	September 2018

[17]	S.Saru	Pima Indian diabetes database	768	N/A	Decision Trees	94.44%	To predict diabetes	April 2019
[18]	Protap Kumar Saha	Pima Indian diabetes database	768	N/A	Neural Network	80.4%	To predict diabetes	December 2019
[19]	Roshan Birjais	Pima Indian diabetes database	768	Sensitivity, Specificity	Gradient boosting	86%	diagnosis of diabetes	28 August 2019
[20]	An Dinh	National Health and Nutrition Examination Survey (NHANES) dataset	5000	Precision, Recall and F1 score	XGBoost	86.2% 73.7%	To predict diabetes and Prediabetic	2019
[1]	Hang Lai	CPCSSN	13,309	N/A	Gradient Boosting Machine (GBM)	84.7%	To predict diabetes	2019
[21]	Y. Jeevan Nagendra Kumar	Pima Indian diabetes database	768	Precision Recall F1 score	KNN	*85%*	To predict diabetes	2019
[22]	Kucharlapati Manoj Varma	Pima Indian diabetes database	768	sensitivity, specificity, Precision and F1 Score	Decision Tree	74.63%	To compare the performance analysis	June 2019
[23]	Mitesh Warke	National Institute of Diabetes and Digestive and Kidney Diseases	1500	N/A	Naive Bayes	72%	To predict diabetes	Mar 2019
[24]		Pima Indians Diabete	768	sensitivity, specificity,	Deep Neural Network	98.35 %	To predict diabetes	

	Safial Islam Ayon	s Dataset		F1-score and MCC				08 March 2019
[25]	Ms. Komal Patil	Pima Indians Diabete s Dataset	768	precision, recall, F1-score and AUC	Stacking Classifier and ADA Boost.	80.08%	To predict diabetes	Novem ber-2019
[26]	Amani Yahyaoui	Pima Indians Diabete s Dataset	768	precision, recall	*Random Forest*	83.67%	To predict diabetes	2019
[27]	Priyanka Sonar	Pima Indians Diabete s Dataset	768	precision, recall, F1-score	Decision Tree	85%	To predict diabetes	2019
[28]	Talha Mahboob Alam	Pima Indians Diabete s Dataset	768	TP, FN, FP, TN	Artificial neural network (ANN)	75.7%	To predict diabetes	09 July 2019
[5]	Md. Maniruzza man	Nationa l Health and Nutritio n Examin ation Survey (NHA NES).	6561	AUC	combination of LR and RF-based classifier	94.25%	To predict diabetes	2020
[2]	KM Jyoti Rani	Pima Indians Diabete s Dataset	768	N/A	Decision Tree	98%	To predict diabetes	July-30-2020
[29]	Aeshah Saad Alanazi	collecte d from diabetic patients at the Securit y Force Primar y Health Care in Tabuk, Saudi Arabia.	non	precision, recall	Random Forest	98%	To predict diabetes	*Sep. 2020*
[30]	MD. KAMRUL HASAN	Pima Indians Diabete s Dataset	768	Sensitivity, specificity, false omission rate, diagnostic odds ratio, and AUC	ensembling classifier (AdaBoost, XGBoost)	95%	To predict diabetes	April 23, 2020,

[31]	S. M. Mahedy Hasan1,	Pima Indians Diabetes Dataset	768	precision, recall, F1-score, ROC-AUC	Tree-Based algorithms, to improve the performance Extra Tree (ET) algorithm as a base estimator of the AdaBoost classifier	90.5%	To predict diabetes	29 November 2020,
[32]	Gaurav Tripathi	Pima Indians Diabetes Dataset	768	Precision, Recall, Specificity and Fscore	Random Forest (RF)	87.66 %	To predict diabetes	2020
[33]	Jingyu Xue	Sylhet Diabetes Hospital in Sylhet	520	TP, FN, FP, TN	SVM	96.54%	To predict diabetes	2020
[34]	Jobeda Jamal Khanam	Pima Indians Diabetes Dataset	768	Precision Recall F-measure	Neural Network (NN) model	88.6%	To predict diabetes	20 February 2021
[35]	Haohui Lu	CBHS health funds company in Australia	124,000	TP, FN, FP, TN precision, recall, F1-score	random forest	83.98	To predict T2DM	10 June 2021
[36]	Fayroza Alaa Khaleel	Pima Indians Diabetes Dataset	768	precision, recall, F1-score	Logistic Regression (LR),	94%	To predict diabetes	2021
[37]	Henock M. Deberneh and Intaek Kim	Hanaro Medical foundation in Seoul, South Korea	535,169	Precision, Recall, F1-score, MCC and KC	RF, SVM, CIM, Soft voting	73%	To predict T2DM	23 March 2021
[38]	Saloni Kumari	Pima Indians Diabetes Dataset	768	precision, recall, F1-score	ensemble soft voting classifier (random forest, logistic regression, and Naive Bayes)	79.08%	To predict diabetes	30 January 2021
[39]	Muhammad Mazhar Bukhari	Pima Indians Diabete	768	MSE	artificial back propagation scaled conjugate	93%	To predict diabetes	22 April 2021

		s Dataset			gradient neural network (ABPSCGNN)			
[40]	Nazin Ahmed	Pima Indians Diabete s Dataset - Tigga and Garg (2020)	768-950	precision, recall, F1-score, ROC	Decision tree (DT), Random Forest (RF), Support Vector Machine (SVM).	80.26 % RF, SVM 96.81% DT, RF	To predict diabetes	7 Decem ber 2021
[41]	Yaqi Tan	8787 desensi tization physica l examin ation data from Qingda o CDC.	520	average prediction time, precision, sensitivity, specificity, and F1-score	GA-stacking ensemble learning model (stacking based on CNN and SVM) (GA for the Feature Selection in Classifiers Based on DT)	98.71%	To predict diabetes	10 Januar y 2022
[42]	Md. Mehedi Hassan	Early Stage Diabete s Risk Predicti on dataset	520	Precision, Recall, MSE, F1-Score, TP, FN, FP, TN	RF	99.03%	To predict diabetes	21 Septem ber 2022
[43]	Isfafuzza man Tasin	Pima Indians Diabete s Dataset - RTML private dataset	768 - 203	Precision, Recall, F1 Score and AUC	XGBoost	81%	To predict diabetes	2022
[44]	Rasool Jader	Collect ed from Kurdist an region laborat ories	1012	Precision, Recall,	max voting method	92%	To Diagnosis Gestational Diabetes	22 Octobe r 2022
[45]	Yifan Qin	1999–2020 NHAN ES databas e	17,833	sensitivity, specificity, precision, F1 score, and (ROC)	CATBoost	82.1%	To predict diabetes	15 Novem ber 2022
[46]	Shahid Mohamma d Ganie	Collect ed data from	1939	specificity, precision, recall, F1	bagged decision tree	99.41%	To predict T2DM	9 August 2022

		different hospitals		score and MCR				
[47]	USAMA AHMED	*UCI Machine Learning Repository.*	520	specificity, sensitivity, precision, and F1 score	Support Vector Machine (SVM) and Articial Neural Network (ANN)-> fuzzy model (fused ML)	94.87	To predict diabetes	January 11, 2022,
[48]	Zaigham Mushtaq	Pima Indians Diabetes Dataset	768	sensitivity, specificity, and ROC	voting (Nave Bayes, RF, and Gradient Boosting)	81.7	To predict diabetes	19 March 2022
[49]	Chollette C. Olisah	Pima Indians Diabetes Dataset and LMCH diabetes datasets	768	precision, sensitivity, F1- score	twice-growth deep neural network (2GDNN) model	97.25 %, 97.33%	To predict diabetes	22 March 2022
[50]	Michael Onyema Edeh1	Pima Indians Diabetes Dataset - Frankfurt Hospital		Recall score F1 score	Random forest, SVM	83.1% svm 97.6% rf	To predict diabetes	31 March 2022
[51]	Umm e Laila	Pima Indians Diabetes Dataset	768	precision, recall and f1-score.	Random Forest	97%	To predict diabetes	13 July 2022
[52]	T. R. Mahesh	Kaggle dataset	693	Precision Recall *F*1_score	EL based on Bayesian networks and radial basis functions	97.11%	To predict diabetes	14 July 2022
[53]	B. Shamreen Ahamed	Pima Indians Diabetes Dataset - DMS	768, 1110	F1-score, precision, recall, sensitivity, and specificity	LGBM	98.99% for DMS 92.5% for pima	To predict diabetes	19 September 2022
[3]	B. Shamreen Ahamed	Pima Indians Diabete	768, 734	precision, recall, specificity,	Light Gradient	95.20%	To predict diabetes and	30 Decem

		s Dataset and a clinical survey dataset.		and sensitivity.	Boosting Machine		type classification	ber 2022
[54]	Muhammad Zarar	Pima Indian diabetes dataset and the Kaggle diabetes data set	768, 1316	sensitivity, specificity, and area under the receiver operating characteristic curve	Artificial Neural Network	98.8%	To predict diabetes	July 12th, 2023
[6]	Subhash Chandra Gupta	Pima Indians Diabetes Dataset	768	F1score	Random Forest	88.61%	To predict diabetes	2023
[55]	Alain Hennebelle	Pima Indians Diabetes Dataset - Sylhet	768 - 520	precision, recall and f1-score.	Random Forest	78.27% PIDD 97.23% Sylhet	To predict T2DM	2023
[4]	Chukwuebuka Joseph Ejiyi	Pima Indians Diabetes Dataset	768	F1 score	Xgboost and Adaboost	94.67%	To predict diabetes	March 2023
[56]	Burçin Kurt	Collected data in Turkey	489	precision, recall, F1-score, sensitivity	RNN-LSTM with Bayesian optimization	95%	To Diagnosis Gestational Diabetes	27 February 2023

References

1. H. Lai, H. Huang, K. Keshavjee, A. Guergachi, and X. Gao, "Predictive models for diabetes mellitus using machine learning techniques," *BMC endocrine disorders,* vol. 19, pp. 1-9, 2019.
2. K. J. Rani, "Diabetes prediction using machine learning," *International Journal of Scientific Research in Computer Science Engineering and Information Technology,* vol. 6, pp. 294-305, 2020.
3. B. S. Ahamed, M. S. Arya, S. Sangeetha, and N. V. Auxilia Osvin, "Diabetes Mellitus Disease Prediction and Type Classification Involving Predictive Modeling Using Machine Learning Techniques and Classifiers," *Applied Computational Intelligence and Soft Computing,* vol. 2022, 2022.
4. C. J. Ejiyi *et al.*, "A robust predictive diagnosis model for diabetes mellitus using Shapley-incorporated machine learning algorithms," *Healthcare Analytics,* vol. 3, p. 100166, 2023.

5. M. Maniruzzaman, M. J. Rahman, B. Ahammed, and M. M. Abedin, "Classification and prediction of diabetes disease using machine learning paradigm," *Health information science and systems,* vol. 8, pp. 1-14, 2020.
6. S. C. Gupta and N. Goel, "Predictive Modeling and Analytics for Diabetes using Hyperparameter tuned Machine Learning Techniques," *Procedia Computer Science,* vol. 218, pp. 1257-1269, 2023.
7. I. Tasin, T. U. Nabil, S. Islam, and R. Khan, "Diabetes prediction using machine learning and explainable AI techniques," *Healthcare Technology Letters,* vol. 10, no. 1-2, pp. 1-10, 2023.
8. M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *International journal of surgery,* vol. 88, p. 105906, 2021.
9. H. M. Qureshi and M. Qureshi, "Diabetes Type1 and Type2 Classification Using Machine Learning Technique," *International Journal of Computer Science and Information Security (IJCSIS),* vol. 15, no. 6, 2017.
10. N. Yuvaraj and K. SriPreethaa, "Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster," *Cluster Computing,* vol. 22, no. Suppl 1, pp. 1-9, 2019.
11. F. G. Woldemichael and S. Menaria, "Prediction of diabetes using data mining techniques," in *2018 2nd international conference on trends in electronics and informatics (ICOEI)*, 2018: IEEE, pp. 414-418.
12. S. Srivastava, L. Sharma, V. Sharma, A. Kumar, and H. Darbari, "Prediction of diabetes using artificial neural network approach," in *Engineering Vibration, Communication and Information Processing: ICoEVCI 2018, India*, 2019: Springer, pp. 679-687.
13. Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in genetics,* vol. 9, p. 515, 2018.
14. A. Mir and S. N. Dhage, "Diabetes disease prediction using machine learning on big data of healthcare," in *2018 fourth international conference on computing communication control and automation (ICCUBEA)*, 2018: IEEE, pp. 1-6.
15. D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia computer science,* vol. 132, pp. 1578-1585, 2018.
16. M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, "Prediction of diabetes using machine learning algorithms in healthcare," in *2018 24th international conference on automation and computing (ICAC)*, 2018: IEEE, pp. 1-6.
17. S. Saru and S. Subashree, "Analysis and prediction of diabetes using machine learning," *International journal of emerging technology and innovative engineering,* vol. 5, no. 4, 2019.
18. P. K. Saha, N. S. Patwary, and I. Ahmed, "A widespread study of diabetes prediction using several machine learning techniques," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019: IEEE, pp. 1-5.

19. R. Birjais, A. K. Mourya, R. Chauhan, and H. Kaur, "Prediction and diagnosis of future diabetes risk: a machine learning approach," *SN Applied Sciences,* vol. 1, pp. 1-8, 2019.
20. A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC medical informatics and decision making,* vol. 19, no. 1, pp. 1-15, 2019.
21. Y. J. N. Kumar, N. K. Shalini, P. Abhilash, K. Sandeep, and D. Indira, "Prediction of diabetes using machine learning," *International Journal of Innovative Technology and Exploring Engineering,* vol. 8, no. 7, pp. 2547-2551, 2019.
22. K. M. Varma and D. B. Panda, "Comparative analysis of Predicting Diabetes Using Machine Learning Techniques," *J. Emerg. Technol. Innov. Res,* vol. 6, pp. 522-530, 2019.
23. M. Warke, V. Kumar, S. Tarale, P. Galgat, and D. Chaudhari, "Diabetes diagnosis using machine learning algorithms," *Diabetes,* vol. 6, no. 03, pp. 1470-1476, 2019.
24. S. I. Ayon and M. M. Islam, "Diabetes prediction: a deep learning approach," *International Journal of Information Engineering and Electronic Business,* vol. 12, no. 2, p. 21, 2019.
25. M. K. Patil, S. Sawarkar, and M. S. Narwane, "Designing a model to detect diabetes using machine learning," *Int. J. Eng. Res. Technol,* vol. 8, no. 11, pp. 333-340, 2019.
26. A. Yahyaoui, A. Jamil, J. Rasheed, and M. Yesiltepe, "A decision support system for diabetes prediction using machine learning and deep learning techniques," in *2019 1st International informatics and software engineering conference (UBMYK)*, 2019: IEEE, pp. 1-4.
27. P. Sonar and K. JayaMalini, "Diabetes prediction using different machine learning approaches," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 2019: IEEE, pp. 367-371.
28. T. M. Alam *et al.*, "A model for early prediction of diabetes," *Informatics in Medicine Unlocked,* vol. 16, p. 100204, 2019.
29. A. S. Alanazi and M. A. Mezher, "Using machine learning algorithms for prediction of diabetes mellitus," in *2020 international conference on computing and information technology (ICCIT-1441)*, 2020: IEEE, pp. 1-3.
30. M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access,* vol. 8, pp. 76516-76531, 2020.
31. S. M. Hasan, M. F. Rabbi, A. I. Champa, and M. A. Zaman, "An Effective Diabetes Prediction System Using Machine Learning Techniques," in *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, 2020: IEEE, pp. 23-28.
32. G. Tripathi and R. Kumar, "Early prediction of diabetes mellitus using machine learning," in *2020 8th international conference on reliability, Infocom technologies and optimization (trends and future directions)(ICRITO)*, 2020: IEEE, pp. 1009-1014.

33. J. Xue, F. Min, and F. Ma, "Research on diabetes prediction method based on machine learning," in *Journal of Physics: Conference Series*, 2020, vol. 1684, no. 1: IOP Publishing, p. 012062.
34. J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *Ict Express,* vol. 7, no. 4, pp. 432-439, 2021.
35. H. Lu, S. Uddin, F. Hajati, M. A. Moni, and M. Khushi, "A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus," *Applied Intelligence,* vol. 52, no. 3, pp. 2411-2422, 2022.
36. F. A. Khaleel and A. M. Al-Bakry, "Diagnosis of diabetes using machine learning algorithms," *Materials Today: Proceedings,* vol. 80, pp. 3200-3203, 2023.
37. H. M. Deberneh and I. Kim, "Prediction of type 2 diabetes based on machine learning algorithm," *International journal of environmental research and public health,* vol. 18, no. 6, p. 3317, 2021.
38. S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *International Journal of Cognitive Computing in Engineering,* vol. 2, pp. 40-46, 2021.
39. M. M. Bukhari, B. F. Alkhamees, S. Hussain, A. Gumaei, A. Assiri, and S. S. Ullah, "An improved artificial neural network model for effective diabetes prediction," *Complexity,* vol. 2021, pp. 1-10, 2021.
40. N. Ahmed *et al.*, "Machine learning based diabetes prediction and development of smart web application," *International Journal of Cognitive Computing in Engineering,* vol. 2, pp. 229-241, 2021.
41. Y. Tan, H. Chen, J. Zhang, R. Tang, and P. Liu, "Early risk prediction of diabetes based on GA-stacking," *Applied Sciences,* vol. 12, no. 2, p. 632, 2022.
42. S. M. Md. Mehedi Hassan, Farhana Yasmin,, "An unsupervised cluster-based feature grouping model for early diabetes detection," *Healthcare Analytics,* vol. 2, 2022.
43. T. U. N. Isfafuzzaman Tasin, Sanjida Islam, Riasat Khan, "Diabetes prediction using machine learning and explainable AI techniques," *The Institution of Engineering and Technology,* 2022.
44. R. Jader and S. Aminifar, "Predictive Model for Diagnosis of Gestational Diabetes in the Kurdistan Region by a Combination of Clustering and Classification Algorithms: An Ensemble Approach," *Applied Computational Intelligence and Soft Computing,* vol. 2022, 2022.
45. Y. Qin *et al.*, "Machine Learning Models for Data-Driven Prediction of Diabetes by Lifestyle Type," *International Journal of Environmental Research and Public Health,* vol. 19, no. 22, p. 15027, 2022.
46. S. M. Ganie and M. B. Malik, "An ensemble machine learning approach for predicting type-II diabetes mellitus based on lifestyle indicators," *Healthcare Analytics,* vol. 2, p. 100092, 2022.
47. U. Ahmed *et al.*, "Prediction of diabetes empowered with fused machine learning," *IEEE Access,* vol. 10, pp. 8529-8538, 2022.
48. Z. Mushtaq, M. F. Ramzan, S. Ali, S. Baseer, A. Samad, and M. Husnain, "Voting classification-based diabetes mellitus prediction using hypertuned

machine-learning techniques," *Mobile Information Systems,* vol. 2022, pp. 1-16, 2022.

49. C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Computer Methods and Programs in Biomedicine,* vol. 220, p. 106773, 2022.
50. M. O. Edeh *et al.*, "A classification algorithm-based hybrid diabetes prediction model," *Frontiers in Public Health,* vol. 10, p. 829519, 2022.
51. U. e. Laila, K. Mahboob, A. W. Khan, F. Khan, and W. Taekeun, "An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study," *Sensors,* vol. 22, no. 14, p. 5247, 2022.
52. T. Mahesh *et al.*, "Blended ensemble learning prediction model for strengthening diagnosis and treatment of chronic diabetes disease," *Computational Intelligence and Neuroscience,* vol. 2022, 2022.
53. B. S. Ahamed, M. S. Arya, and A. O. V. Nancy, "Diabetes Mellitus Disease Prediction Using Machine Learning Classifiers with Oversampling and Feature Augmentation," *Advances in Human-Computer Interaction,* 2022.
54. M. Zarar and Y. Wang, "Early Stage Diabetes Prediction by Approach Using Machine Learning Techniques," 2023.
55. A. Hennebelle, H. Materwala, and L. Ismail, "HealthEdge: a machine learning-based smart healthcare framework for prediction of type 2 diabetes in an integrated IoT, edge, and cloud computing system," *Procedia Computer Science,* vol. 220, pp. 331-338, 2023.
56. B. Kurt *et al.*, "Prediction of gestational diabetes using deep learning and Bayesian optimization and traditional machine learning techniques," *Medical & Biological Engineering & Computing,* pp. 1-12, 2023.