JCHR

Journal of Community Health Research 2026; 15: 18-27.

Comparison of Artificial Neural Network and Decision Tree Model Algorithms in Predicting Type 2 Diabetes Classification Status

Masoud Amiri [1], Somayeh Hessam [1*], Shaghayegh Vahdat [1], Shahram Tofighi [2, 3]

1. Department of Health Service Administration, ST.C., Islamic Azad University, Tehran, Iran
2. Department of Future Studies and Theory Building, Iranian Academy of Medical Sciences, Tehran, Iran
3. National Center for Health Insurance Research, Tehran, Iran

ARTICLE INFO

Original Article

Received: 1 Mar 2026

Accepted: 5 May 2026

OPEN ACCESS

Corresponding Author:
Somayeh Hessam
Shessam@iau.ac.ir

ABSTRACT

Background: Diabetes is a metabolic disorder in the body. Using data mining techniques is useful for predicting diabetes, so the aim of this study was to predict diabetes status using artificial neural network and decision tree models.

Methods: This study was descriptive and based on secondary data. Data from 4820 individuals were also analyzed. In this study, the performance of two decision tree models and artificial neural networks was compared. Data was randomly divided into three parts, 70% as training, 20% as validation, and 10% as testing. Various criteria such as accuracy, precision, specificity, sensitivity, ROC-AUC curve, and F1-Score were used to evaluate the models, and finally, the best algorithm for predicting diabetes was identified.

Results: The decision tree and artificial neural network models were obtained with 97% Precision and 97% accuracy and 96% Precision and 96% accuracy, respectively. The area under the ROC curve in the artificial neural network model (95%) was higher in the training and testing sets than the decision tree model (92%).

Conclusion: Although the accuracy of the decision tree model in predicting diabetes status was slightly higher than that of the artificial neural network model, the area under the curve (AUC) of the neural network was higher, and therefore, both models performed well. According to these two models, the variables of fasting blood sugar, systolic blood pressure, and age were effective variables in predicting diabetes status.

Keywords: Prediction, Type 2 Diabetes, Machine Learning, Data Mining

How to cite this paper:
Amiri M, Hessam S, Vahdat Sh, Tofighi Sh. Comparison of Artificial Neural Network and Decision Tree Model Algorithms in Predicting Type 2 Diabetes Classification Status. J Community Health Research 2026; 15: 18-27.

[DOI: 10.18502/jchr.v15i3.22092]

Introduction

Type 2 diabetes is the most common metabolic disorder in the world, and several factors, including an unbalanced diet, irregular lifestyle, complex genetic structure, obesity, etc., play a role in the development of diabetes (1). Most people with diabetes are unaware of their disease and have no symptoms, but the most important characteristics and symptoms of diabetes can include increased thirst, frequent urination, and overeating (2). Delayed diagnosis and prediction of diabetes and, as a result, inadequate blood sugar control significantly increases the risk of cardiovascular complications, renal failure, eye complications, and lower limb involvement (3). On average, 10% of the world's population has type 2 diabetes. According to the research on the prediction of the prevalence of diabetes for the years 2030 and 2045, it is estimated that people with diabetes will increase to 10.2% of the world's adult population (578 million people) by 2030 and to 10.9% (700 million people) by 2045(4). The prevalence of diabetes in Iran is 14.2%, which is higher than the global average. There are about 8 million diabetic patients in Iran, which is higher than international estimates (5). One of the effective measures to prevent this disease is to predict the disease in healthy susceptible individuals (6). First-degree relatives of type 2 diabetic patients are at higher risk of developing diabetes (7). In addition, the economic impact of this disease is significant. The global economic burden of diabetes is expected to reach approximately $802 billion by 2040 (8). In Iran, a large amount of money is spent annually on diabetes-related care and disability. Since prediction precedes treatment, accurate and reliable techniques must be used to ensure that predictions are accurate and reliable (9). One of these methods is the use of machine learning and data mining algorithms. Machine learning (ML) techniques are valuable and effective tools for identifying individuals at risk of developing and identifying risk factors associated with T2DM. Models such as decision trees and neural networks have high accuracy for identifying and predicting diabetes. (10-15). Various studies have been conducted worldwide and in Iran, some of which have shown the superiority of decision tree algorithms and artificial neural networks together, while other studies have shown the superiority of decision tree algorithms and neural networks separately with different accuracy in predicting diabetes (16, 17). Therefore, the aim of this study is the comparison of artificial neural network and decision tree model algorithms in predicting type 2 diabetes classification status.

Methods

This descriptive study was based on secondary data which was conducted based on 5 basic and important steps. These five steps include data collection, data preprocessing and feature extraction, selection of machine learning models, data segmentation, evaluation of models with different criteria, and finally, selection of the best model to predict type 2 diabetes. The study population included people who participated in the diabetes screening campaign and were extracted from the Ministry of Health's Sib system. In this study, 4820 people who had complete information were entered into the model. Of these, 507 were diabetic, 822 were prediabetic, and 3491 were healthy. The total number of the variables extracted from the Ministry of Health's SIB system from the Diabetes Screening Campaign was 14. In order to determine the factors affecting the incidence of diabetes, 40 domestic and foreign articles in English and Persian were reviewed in a 10-year period from 2015 to 2025. In order to determine which variables had the most risk factors and effects on diabetes in the articles, age, BMI, fasting blood sugar, family history of type 1 diabetes, inactivity, history of high blood pressure, systolic blood pressure, and smoking history were included, which were entered into the model as predictor variables. These parameters were identified for designing predictive models through a systematic review study, but four other variables were excluded from this study due to their least amount of importance on diabetes, which included family history of cardiovascular disease, place of residence, height, and weight.

[DOI: 10.18502/jchr.v15i3.22092]

Finally, the total number of variables in this study was 10. The study's inclusion criteria included being native to the city, having complete data, and all screened individuals being 30 years of age or older. In this study, the fasting blood sugar variable was used only as a dependent variable, which was classified into three categories according to the instructions, including blood sugar less than 100 mg/dl as healthy, blood sugar between 100 and 126 mg/dl in the prediabetes stage, and above 126 mg/dl as type 2 diabetes. It is emphasized that fasting blood sugar was not used as an input to the model (independent variable). Data preprocessing was performed to clean the data, remove or fill in missing values, normalize, standardize, and remove duplicate data or noise. In this study, due to data collection method, no missing data was identified in these 10 variables, and all necessary data were extracted. This study aims to predict diabetes based on some effective variables and with the help of machine learning algorithms. Therefore, artificial neural network (ANN) and decision tree (DT) models have been used to analyze the data. The neural network implementation in the present study was carried out with the MLP architecture and TensorFlow and Keras libraries in the Python environment. Moreover, the input layer consisted of 10 neurons to receive independent variables (except fasting blood sugar, which was considered the output) and the output layer consisted of 3 neurons with the Softmax function and One-Hot coding for healthy, prediabetic, and diabetic classes. The selected architecture has two hidden layers with 128 and 64 neurons and the ReLU function. Also, the training process was performed with the Adam optimizer (learning rate 0.001) and the Categorical Cross-Entropy cost function. In order to control overfitting, Early Stopping with Patience equal to 10 and validation error monitoring were used, and the Batch Size and maximum Epochs values were considered to be 32 and 200, respectively. The meta-parameters (number of neurons, learning rate) were adjusted with Grid Search and 5-layer validation on the training data. Decision tree model was used as well to classify the diabetic status of individuals. The implementation of the decision tree model in the present study was carried out with the CART algorithm and the Scikit-learn library in the Python environment. The node splitting criterion, Gini Impurity index, and 10 independent variables (except fasting blood sugar, which was the output) were considered as inputs, and three classes of healthy, prediabetic, and diabetic were considered as the model outputs. The meta-parameters were adjusted using Grid Search and 5-Fold CV on the training data. Furthermore, the range for maximum depth (max_depth) was 3 to 15; for minimum samples to split the node (min_samples_split), 2 to 50; and for minimum samples in the leaf (min_samples_leaf), 5 to 20 were defined; their optimal values were obtained based on the lowest validation error, respectively, as 5, 20, and 10. To control overfitting, cost-complexity pruning was applied with a ccp_alpha coefficient of 0.01. The split strategy was set to best, and max_features was set to the square root of the total variables (sqrt) to avoid excessive complexity. In using both models, data was randomly divided into three parts, respectively, 70% of the data was used as training samples to build the machine learning models used; 20% of the data was used as a validation set to adjust hyperparameters (learning rate, number of layers in neural networks, etc.) and prevent overfitting (over-learning of training data), and 10% of the data was used as a test set for final evaluation with indicators. This division was done based on common ratios in data science to ensure that the model is able to predict new data (which are in the test set) well. In the model parameter tuning section, the validation method is used to better evaluate the parameters and improve the accuracy and other metrics of the model. To evaluate the quality of artificial neural network and decision tree models, the criteria of precision, accuracy, sensitivity, specificity, roc-auc curve and F1-score were used. The software version used Python, Jupyter Notebook, and various libraries to implement and analyze the algorithms. In this study, all the codes were written and executed in Jupyter Notebook. In addition, the initial processing and implementation of machine learning algorithms was done with the Scikit-learn library. Deep learning models were built and trained

[DOI: 10.18502/jchr.v15i3.22092]

using TensorFlow and Keras. The Python language only played the role of the runtime environment and library coordinator, and finally, the best algorithm with high accuracy as well as the effective variables in diabetes were identified to predict the diabetic status.

Results

Table 1 shows the information related to 4820 people who participated in the diabetes screening campaign. Of these, 507 were diabetic, 822 were prediabetic, and 3491 were healthy.

Table 1. classification of people

Diabetes status	Total (%)	Male (%)	Female (%)
Diabetes	507(%10.51)	232 (% 45.8)	275 (% 54.2)
Prediabetes	822 (%17.04)	428 (% 52.1)	394 (% 47.9)
Healthy	391(%72.45)	1839 (% 52.7)	1652 (% 47.3)
Total	4820 (%100)	2499 (% 100)	2321 (% 48.2)

According to Table 1, the total number of samples was 4820 people, of whom 507 were diabetic patients, which is equivalent to 10.51% of the total sample, 822 were prediabetic, and the rest were healthy, which accounts for 72.3% of the total sample. These people entered this study in three groups: healthy, prediabetic, and diabetic.

The importance of variables in the decision tree model was measured based on the Gini Importance criterion. According to this index, fasting blood sugar, systolic blood pressure, and age were identified as the most important predictors; in contrast, gender, history of high blood pressure, physical inactivity, and smoking history had the least effect on predicting diabetes.

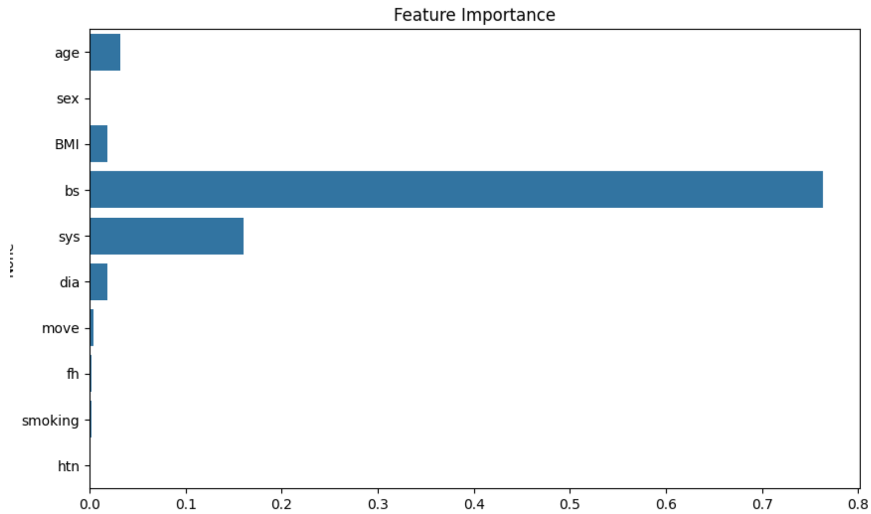


Chart 1. Ranking of diabetes risk factors in the decision tree model

To determine the importance of variables in the neural network model, Permutation Importance method was used. Based on the findings, fasting blood sugar, systolic blood pressure, and age played the most important role in predicting prediabetes, and inactivity, smoking history, and high blood pressure history were given lower priorities.

[DOI: 10.18502/jchr.v15i3.22092]

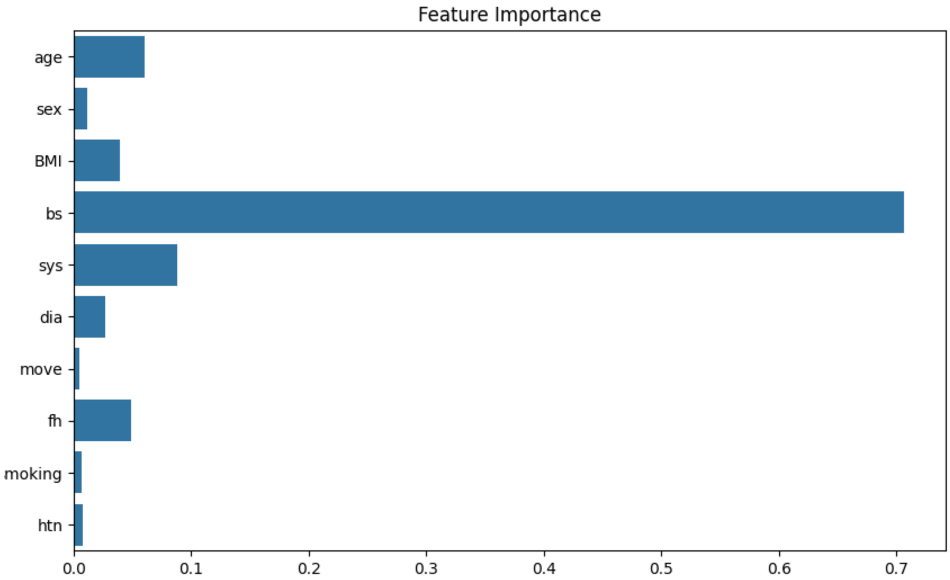


Chart 2. Ranking of diabetes risk factors in the artificial neural network model

According to Table 2, to compare the performance of the models, the performance evaluation indices of the models were performed on the data set. To compare the decision tree and neural network models, the prediction of diabetes status for all sample individuals was obtained with both models.

Table 2. Evaluation criteria for artificial neural network and decision tree models

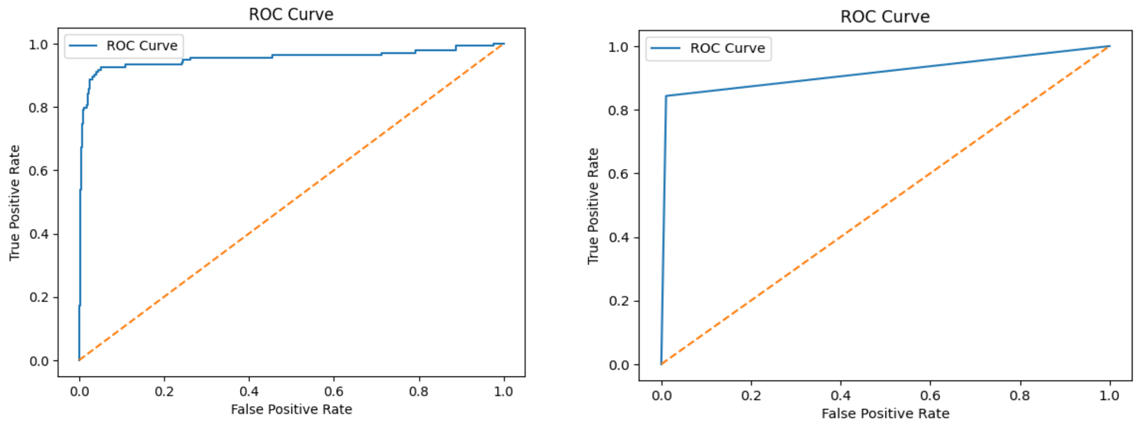
Model used	Amount of training				Amount of testing			
	Precision	Accuracy	Sensitivity	Specificity	Precision	Accuracy	Sensitivity	Specificity
Decision tree	0.9758	0.9754	۰/۹۸۳۲	0.9758	0.97	0.97	0.97	0.97
Artificial neural network	0.9689	0.9682	0.9689	0.9689	۰/۹۶	۰/۹۶	0.97	0.97

According to Table 2, the decision tree model had an accuracy of 97%, and the artificial neural network had an accuracy of 96%. Although the accuracy of the decision tree model in predicting diabetes status was slightly higher than that of the artificial neural network model, both performed well in predicting diabetes status. Also, the accuracy of the decision tree model and the artificial neural network were 97 and 96%, respectively, which despite a very slight difference, both models had the desired performance.

Table 3. Comparison of the area under the Rock curve and F1-Score of artificial neural network and decision tree models

Algorithm used	Amount of training		Amount of testing	
	F1-Score	(AUC-ROC)	F1-Score	(AUC-ROC)
Decision tree	0.9163	0.9756	۰/۹۲	۰/۹۷
Artificial neural network	0.9526	0.9685	۰/۹۵	۰/۹۷

[DOI: 10.18502/jchr.v15i3.22092]



Decision tree   Artificial neural network

Chart 3. Rock curve of artificial neural network and decision tree models

According to Table and Chart 3, the comparison of the areas under the Rock curve in the two artificial neural network and decision tree models showed that these two models differ in terms of the size of the area under the Rock curve in the two test and training sets. In other words, the area under the Rock curve of the artificial neural network model in the training and test sets is greater than that of the decision tree model. The AUC value of the artificial neural network model, which is equal to 0.95, shows that the artificial neural network model has been able to have a significant performance in simulating the state of diabetes.

Furthermore, , the F1-Score criterion for the two models was that the performance of the artificial neural network, and decision tree models on similar test data was the same, but in the training data, the decision tree model showed a better performance.

Discussion

Diabetes is one of the most common metabolic diseases, and due to its high prevalence, access to services that allow for the prediction of diabetes is very practical and effective.

In the study by Meshkouti et al. (18), normalization was performed for 8 variables. The results of this study showed that the support vector machine performed better than the artificial neural network in diabetes classification with an accuracy of 99.6%. In this study, the accuracy of the decision tree model in predicting diabetes was slightly higher than the artificial neural network model, but the area under the curve (AUC) of the neural network was higher, so both models performed well. According to these two models, the variables of fasting blood sugar, systolic blood pressure, and age were effective variables in predicting diabetes. Meshkouti's study, in contrast to this study's results, showed that the support vector machine model performed better than the artificial neural network. While the study by Meshkouti et al. focused on fewer features and a smaller sample size, this study used more data and more advanced techniques including validation and meta-parameter tuning to improve the accuracy of predicting diabetes status, which will increase the performance of the model and was an advantage of this study. According to these two studies, in both models, fasting blood sugar, systolic blood pressure, and age were effective variables in predicting diabetes status. These three variables can be given special attention as key features in the healthcare system to prevent the occurrence of diabetes, especially in families with a history of diabetes.

In the study by Abeer El-Sayyid et al. (19) in 2026, 9 machine learning algorithms including ANN, LR, NB, DT, AB, RF, GB, ET and KNN were used for PIMA database on 34 variables. Finally, the result of the study showed that the artificial neural network model with an accuracy of 99.98% had the best performance among the 9 models entered, and then, the decision tree with an accuracy of 99.75% was used as the next model for prediction, which was in line with this study and both the decision tree

[DOI: 10.18502/jchr.v15i3.22092]

and neural network models performed well with a slight difference in accuracy and AUC. However, the accuracy of Abeer El-Sayyid's study was higher than this study, which could be due to the number of parameters and features (34 versus 10), the type of data (laboratory in behavioral and biographical articles), and the different types of algorithms. But, one strength of this study was the sample size of 4820 people compared to 768 people. Age and history of blood pressure were also identified as the most important risk factors for diabetes., only the risk factor of age in the present study was similar to this study, which emphasized the consistency of the findings of machine learning models with basic medical knowledge. The age variable was the most important predictor of diabetes in these two studies, which can provide valuable information for doctors and patients to take preventive measures to prevent the occurrence of diabetes; the other variables were of less importance in developing diabetes.

In another study by Salem Alzboon et al (20) in 2025, the prediction and management of diabetes using machine learning approaches was carried out. In this study, PID Dataset data and seven types of machine learning models including LR, DT, RF, S VM, NB, GB, and NN were used. The results showed that the neural network had the highest accuracy among the models, and then, the decision tree had the best performance in predicting diabetes. The results of this study were almost similar to our study. Considering that there was a very slight difference between the two models, it could be due to the nature of the data. In our study, the decision tree had a higher accuracy, but in the Salem Alzboon study, the neural network had a higher accuracy. In Salem Alzboon, there were 6 algorithms, but in our study, 2 algorithms were used. The diversity of the algorithms in our study was low, but the sample size in the two studies was also high, which indicates generalizability to other populations.

In a study conducted by Ahamed, B et al. (21) , PIMA data and 6 machine learning methods and models including DT, SVM, KNN, RBF, ANN, and MDR were used. The results of the study showed that decision tree and SVM models had the highest accuracy and precision in predicting diabetes, which was similar to the results of this study in terms of model accuracy performance, but the diversity of the study models was less than that of Ahamed's study, while the diversity of models in Ahamed's study was six, which allows for more comparison between models; it has greater generalizability than this study in terms of the multiplicity of algorithms.

In the study by Reza, M et al. (22) , normalization was performed using the 5-fold cross-validation method, and data was divided into training and testing sets. Among the different machine learning models, the neural network model achieved the best performance among all models with an accuracy of 95.50%, which confirms the findings of this study for both the decision tree and neural network models. In the study by Reza, M, a simple segmentation method was used, while in the present study, an advanced validation method was used to randomly divide the data into three modes: 70 % training, 20 % validation, and 10 % test set, which is important for the validity and reliability of the results of this study.

Moreover, the findings of the study by Habibi S and Huang Y showed that the decision tree model performed better in predicting diabetes than other machine learning models, which is similar and in line with the results of this study (23,24). However, the number of variables in these studies was less than the present study. This study showed that by considering more complete data from medical history, the prediction accuracy of the models improves. Krishnamoorthi used various machine learning methods including DT, NN, KNN, RF, NB, AB, LR, and SVM in his study, and showed that the logistic regression model performed better in predicting diabetes than the other models, which in this study indicates the higher ability of other techniques compared to decision trees and neural networks; this was contrary to the results of the present study (25). In Krishnamoorthi's study, more algorithms (8 vs. 2) were used and entered into the model compared with the present study, which allows for more comparisons. While the methods for evaluating models (accuracy, precision, etc.) were more diverse, they showed different performance in the training and testing sets.

In a study by Abolfazl Kazemi et al. (26) 3 machine learning algorithm models including decision tree, neural network and Bayesian network were used for 500 people. The findings showed that the decision tree model has the highest accuracy with 95.68%, and the Bayesian network has the lowest accuracy in diagnosing diabetic patients; is the results were consistent with the findings of the present study. Among the stated features, blood lipid variables, age and sugar level have the greatest impact on predicting the occurrence of diabetes, which were of the highest significance in the common feature of the 2 variables of fasting blood sugar and age in both studies. These two variables have been identified in most studies as the most important risk factors for diabetes, which can be considered and planned by policymakers in the field of health and treatment; the rest of the variables were less important. One of the innovations of this research was the use of a higher sample size, while in Kazemi's study, the sample size included 500 people. The smaller sample size in this study led to more limited features and less validity and generalizability to the population compared to the current study.

In the study by Mohammad Maydanchi et al. (27), 9 machine learning classification models were used in this study, including XGBoost, Gradient Boosting, AdaBoost, Logistic Regression, Decision Tree, KNN, Perceptron, Random Forest, and Naive Bayes. Based on the results , the decision tree showed the highest f1 and AUC scores among the 9 models, which confirmed the findings of the present study regarding the fact that the decision tree and the neural network performed well with a very slight difference. The most important strength of Maydanchi's study over the curent study was the use of 9 algorithms compared to 2 algorithms, but the sample size of this study and the number of features and variables used in this study were larger, which shows greater generalizability and validity.

Similarly, the findings of the studies by Moa Lugner et al. Mi Jin Noh et al. Qing Liu et al. and Hosam El-Sofany et al. did not confirm the results of the present study regarding the fact that the decision tree model, and then, the neural network performed better in predicting diabetes; contrary to the results of the current study, another algorithm performed better in predicting diabetes (28-31). This difference could be due to the difference in the sample size and data, the variables used in each study, and the type of clinical demographic variables selected, which caused a different performance in the models and criteria for evaluating the machine learning algorithm. In addition, in this study, complete evaluation criteria were used to compare two performances, which is one of the advantages of this study, but in the mentioned studies, the evaluation criteria were not mentioned completely and comprehensively.

The findings of the study by Khoula Al Sadi et al. and Ruhollah Kalhor et al. in 2019 were consistent with the findings of the present study (32,33). Given that the sample size examined in these two studies was almost similar to the present study; this agreement could probably be similar in this sample size. The number of algorithms used in the above studies was on average between 2 and 3 models, which was similar to this study.

One of the strengths of this research is that by choosing the right algorithm and adjusting the meta-parameters appropriately, a very accurate and reliable model for predicting diabetes can be created, which will greatly assist physicians and health systems in early identification of patients at risk and timely intervention. The disadvantages of this article are that the sample size is limited to only one city, it only used two machine learning models, and it lacked access to the Ministry of Health's SIB system for data collection.

Conclusion

Although the accuracy of the decision tree model in predicting diabetes status was slightly higher than that of the artificial neural network model, the area under the curve (AUC) of the neural network was higher, and therefore, both models performed well. This study also showed that, based on these two models, the variables of fasting blood sugar, systolic blood pressure, and age among the variables studied had a more effective role in predicting diabetes status and could be used by health policymakers for preventive interventions.

[DOI: 10.18502/jchr.v15i3.22092]

Limitations

Limitations of this study included the dependence of the results on data from a specific population, the lack of external validation of the models, and the possibility of bias due to the clinical variables used.

Acknowledgments

The authors would like to thank all those who helped them carry out the research.

Conflict of Interest

The authors declared no conflict of interest.

Funding

No specific funding was received for the preparation or publication of this editorial

Ethical considerations

This article was derived from a doctoral dissertation at Islamic Azad University, South Tehran Branch, in accordance with the approval of the ethics committee of this university.

Code of ethics

The code of ethics for this research is IR.IAU.VARAMIN.REC.1404.049.

Authors' contributions

The authorship criteria for each author were met according to the recommendations of the International Committee of Medical Journal Editors, were read and approved in the final version.

Open access policy

JCHR does not charge readers and their institution for access to its papers. Full text download of all new and archived papers are free of charge.

Reference

1. N Ahmed, RA, M,M Islam, et al. Machine Learning Based Diabetes Prediction and Development of Smart Web Application. International Journal of Cognitive Computing in Engineering. 2021;2:229–41.
2. Hakeel M. Diabetes prediction machine learning-based diabetes prediction app using random forest algorithm. JATI (Jurnal Mahasiswa Teknik Informatika. 2025;9(1):1370–6.
3. Amri Z, Rodi M, Wathani N, et al. Amri, Z., Rodi, M., Wathani, M. N., & Bagja, A. (2025). Prediksi Diabetes Menggunakan Algoritma K-Nearest (KNN) Teknik SMOTE-ENN. INFOTEK: JURNAL INFORMATIKA DAN TEKNOLOGI Учредители: Universitas Hamzanwadi. 2025; 8(1): 193-204
4. Wee BF, Sivakumar S, Hann Lim k, et al. Diabetes detection based on machine learning and deep learning approaches. Multimedia Tools and Applications. 2024;83(8):24153–85.
5. Nazari B. Investigating the relationship between dietary intake and factors affecting it in patients with type 2 diabetes. Journal of Qom University of Medical Sciences. 2020;14(10):1–13. [Persian]
6. Liu G LY, Hu Y, Zong G, et al. Influence of lifestyle on incident cardiovascular disease and mortality in patients with diabetes mellitus. J AmColl Cardiol. 2018;71(25):2867–73.
7. Liu Y-Q, CT-W, Lee L-C, et al. Use of Machine Learning to Predict the Incidence of Type 2 Diabetes Among Relatively Healthy Adults: A 10-Year Longitudinal Study in Taiwan. Diagnostics. 2025;15(1):72.
8. Khan FA ZK, Al-Rakhami M, Derhab A, et al. Detection and prediction of diabetes using data mining: a comprehensive review. IEEE Access. 2021;9:43711–35.
9. Ali A F, A Saeed. A comparative analysis of machine learning algorithms to build a predictive model for detecting diabetes complications. Informatica. 2021;45(1).
10. Lei T. Diabetes risk assessment: A comparative study of decision trees and ensemble learning models. ITM Web of Conferences Vol 70 EDP Sciences. 2025;70:02020.
11. Benhar H, Idri A, Fernández-Aleman, et al. Data preprocessing for decision making in medical informatics: potential and analysis. In World conference on information systems and technologies Cham: Springer International Publishing. 2018:1208–18.
12. Janghorbani M, Amini M. Artificial Neural Network and Decision Tree Models for Predicting Prediabetes in the 17-Year "Isfahan Diabetes Prevention" Cohort Study. The First International Conference on Statistical Data Analysis. 2023;1. [Persian]
13. Li J HJ, Zheng L, Li X. Application of artificial intelligence in Diabetes education and management: Present status and promising prospect. Frontiers in Public Health 2020;8(173):1–8.

[DOI: 10.18502/jchr.v15i3.22092]

14. M Maniruzzaman, MJR, B Ahammed M, et al. Classification and prediction of diabetes disease using machine learning paradigm. Health Information Science and Systems 2020;8:1–14.
15. Islam MM RH, Shahid MSB, Akhter A, et al. Explainable Machine Learning for Efficient Diabetes Prediction Using Hyperparameter Tuning, SHAP Analysis, Partial Dependency, and LIME. Engineering Reports. 2025;7(1):e13080.
16. Oumoulylte M, Farhaoui Y, El Allaoui A. An efficient prediction system for diabetes disease based on machine learning algorithms. Data and Metadata. 2023;2:173–85.
17. Ranvir Kaur K. Diabetes Prediction Using Machine Learning. Proceedings of Fifth International Conference on Computing and Communication Networks. 2026;9(9).
18. He J BS, Xu J, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. Nature medicine. 2019;25(1):30–6.
19. Sadiq IZ KB, Ibrahim B, Ibrahim M, et al. Data-driven diabetes mellitus prediction and management: a comparative evaluation of decision tree classifier and artificial neural network models along with statistical analysis. Scientific Reports. 2025;15(1):19339.
20. Salem Alzboon M, Alqaraleh, M Subhi, et al. Diabetes prediction and management using machine learning approaches. Data Metadata 2025; 2506.11501
21. Ahamed BS, Arya M, Sangeetha S, et al. Diabetes mellitus disease prediction and type classification involving predictive modeling using machine learning techniques and classifiers. Appl Comput Intell Soft Comput. 2022; 2022(1):7899364
22. MSea R, R Amin, R Yasmin, et al. Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data. Heliyon 2024;10.
23. Habibi S. A study on diabetes type II predictive models applying data mining techniques in expert systems development [Dissertation]. Tehran: Iran University of Medical Sciences; School of Health Management and Information Science. 2015. [Persian]
24. Huang Y MP, Black N, Harper R. Feature selection and classification model construction on type 2 diabetic patients' data. Artif Intell Med. 2007;41(3):251–62.
25. Krishnamoorthi R. A novel diabetes healthcare disease prediction framework using machine learning techniques. J Healthc Eng. 2022:1684017.
26. Kazemi A BH. Provide a Predictive Model to Identify People with Diabetes Using the Decision Tree. Iranian Journal of Diabetes and Metabolism. 2021;21(3):151–64. Persian.
27. Maydanchi M ZM, Mohammadi M, Ziaei A, et al. A comparative analysis of the machine learning methods for predicting diabetes. Journal of Operations Intelligence. 2024;2(1):230–51.
28. Lugner M RA, Helleryd E, Eliasson B. Identifying top ten predictors of type 2 diabetes through machine learning analysis of UK Biobank data. Scientific reports. 2024;14(1):2102.
29. Noh MJ KY. Diabetes Prediction Through Linkage of Causal Discovery and Inference Model with Machine Learning Models. Biomedicines. 2025;13(1):124–38.
30. Liu Q ZQ, He Y, Zou J, et al. Predicting the 2-year risk of progression from prediabetes to diabetes using machine learning among chinese elderly adults. Journal of Personalized Medicine. 2022;12(7):1055.
31. El-Sofany H E-SS, Karam OH, Abd El-Latif YM, et al. A proposed technique using machine learning for the prediction of diabetes disease through a mobile app. International Journal of Intelligent Systems. 2024;1:6688934.
32. Al Sadi K BW. Prediction Model of Type 2 Diabetes Mellitus for Oman Prediabetes Patients Using Artificial Neural Network and Six Machine Learning Classifiers. Applied sciences. 2023;13(2344):1–22.
33. Kalhor R, Mortezagholi A, Naji F, et al. Designing an intelligent system for diagnosing type 2 diabetes using the data mining approach: brief report. Tehran University of Medical Sciences Journal. 2019;76(12):827–31. Persian.

[DOI: 10.18502/jchr.v15i3.22092]