International Journal of Community Medicine and Public Health
Ukoba O et al. Int J Community Med Public Health. 2025 Jun;12(6):2828-2835
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: https://dx.doi.org/10.18203/2394-6040.ijcmph20251734

Systematic Review

A systematic review of machine learning methods for diabetes mellitus prediction and classification in Nigeria

Obatavwe Ukoba[1]*, Ukoba O. Joseph[2], Ochei L. Charles[3], Peter-Kio B. Opirite[4], Anuku E. Omamuyovwi[2], Obi-Ntumeonuwa Matilda[5]

[1]Department of Community Medicine, Delta State University Teaching Hospital, Oghara, Delta State, Nigeria
[2]Department of Computer Science, Faculty of Applied Sciences, Federal Polytechnic Orogun, Delta State, Nigeria
[3]Department of Computer Science, Faculty of Computing, University of Port Harcourt, Rivers State, Nigeria
[4]Department of Human Kinetics, Health and Safety Studies, Faculty of Natural and Applied Sciences, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Nigeria
[5]Department of Community Medicine, Delta State University Teaching Hospital, Oghara, Delta State, Nigeria

Received: 21 April 2025
Revised: 21 May 2025
Accepted: 22 May 2025

***Correspondence:**
Dr. Obatavwe Ukoba,
E-mail: ukobaobatavwe@gmail.com

ABSTRACT

Diabetes mellitus (DM) is a significant public health issue in Nigeria, affecting millions of people. Early detection and management are crucial to prevent severe complications. Traditional diagnostic methods have limitations, prompting the exploration of machine learning (ML) techniques for more accurate and efficient prediction. This review systematically examines existing studies on ML applications for DM prediction and classification in Nigeria. It analyzes the research methodologies, attributes considered, study areas, and performance of different ML algorithms. The findings reveal that while ML holds promise, research is limited in scope, focusing primarily on the northern regions. Supervised learning algorithms like ANN and decision trees have demonstrated promising results for prediction and classification in Nigerian datasets, with logistic regression being a common tool for risk factor analysis. Furthermore, studies often overlook key risk factors prevalent in the southern population. This review highlights the need for future research that considers the southern population and a wider range of risk factors. It further recommends a decision support system to improve the early detection, management, and outcomes of diabetes in remote regions of Nigeria.

Keywords: Machine learning, Diabetes mellitus, Prediction, Classification

INTRODUCTION

Diabetes mellitus (DM) is a chronic metabolic disorder characterized by high blood sugar levels.[1] It has emerged as a significant public health concern in Nigeria, with a growing prevalence rate.[2] The International Diabetes Federation notes that there are about 382 million people living with diabetes in the world and that by 2035, this will almost double, as 592 million.[3] This rise might be narrowed down to certain factors such as age, sex, ethnicity, family history of diabetes, previous gestational diabetes, the age bracket of the populace, sex, marital status, rate of urbanization, overweight, obesity, unhealthy diet, physical inactivity, quantity of sleep, hypertension, fruit and vegetable intake, smoking, high intake of sugar-sweetened beverages, alcohol, education status, unemployment, poverty and many more.[4] Some of these factors have inadvertently resulted in key health challenges around the world.[4] In 2013, 2.3 million deaths resulted from diabetes, however, in 2021 the World Health Organization noted that about 8.5% of adults aged 17 years and older are diabetic, diabetics or its complications accounts for loss of lives every 7 seconds, and 50% of the cases are under 60 years.[3,5] Therefore, early detection and management of diabetes are crucial to prevent severe complications such as heart disease, stroke, blindness, and kidney failure.[6]

Traditional methods for diabetes diagnosis rely on clinical symptoms and blood sugar level tests, however, these methods may not be entirely effective in identifying individuals at risk before the onset of symptoms considering that various factors ranging from biological, lifestyle, socioeconomic, and psychological factors play important roles in the course of diabetes progression.[7] The numerous factors at play in diabetes progression and inter-relations make it difficult to predict the risk of long-term complications early, hence the concept of machine learning.

Machine learning (ML) algorithms can analyze vast amounts of healthcare data to identify patterns and relationships between various risk factors and the development of diabetes. This information can be used to develop predictive models that can estimate the likelihood of an individual developing the disease long before signs and symptoms manifest. Additionally, ML can be used to classify individuals as diabetic or non-diabetic based on their existing clinical data, hence a promising tool to addressing diabetes challenge and also assists in early prediction.

Although studies conducted globally have explored the potentials of machine learning for diabetes prediction and classification, yet little has been done in Nigeria. These studies have employed various algorithms, achieving promising results. Hence, this review aims to examine the application of machine learning methods in the Nigerian context, focusing on: common data sources and challenges associated with data collection in Nigeria, different machine learning algorithms used for diabetes prediction and classification, comparative performance of these algorithms and factors influencing their effectiveness and future directions for research in this field, including potential benefits and challenges.

ML is a technique that allows computers to learn from data without explicit programming.[8] It involves training a model on data and then using it for classification and prediction. ML focuses on extracting information from data using statistical and computational methods. The process typically involves three phases: training, validation, and testing.[8] Its major focus is to extract information automatically from data, by statistical and computational methods. Basically machine learning has three phases in: the training stage, the validation phase and testing.[8] Data mining uses machine learning tools and techniques but lacks decision making feature.[9]

ML methods can be described as having four types: unsupervised ML, semi-supervised ML and supervised machine learning and hybrid machine learning.

Unsupervised machine learning models only input data say (X) with no matching output data say (Y) are presented. The aim of this learning is modelling the fundamental distributions of the data, to learn more on such data. Association and Clustering problems are examples.[10] In unsupervised anomaly detection techniques we don't train the dataset but presume that most instances are normal and only a very small percentage is anomalous. Secondly, it anticipates that anomalous instance varies from normal once.[11] From these assumptions, similar instances appearing frequently are considered normal, while instances that vary significantly from other majority instances are regarded as anomalies.[12]

In unsupervised techniques, the train and test data set are together. Here, detecting anomalies are done by evaluating the fundamental characteristics the dataset has, as there are not normal labels or anomalous labels. The data to undergo training has inputs with no assigned desired outputs. Hence the dataset is unlabelled. The main task here is to seek from the unlabeled data more knowledge.[8] Association rule and clustering problems uses unsupervised learning. Examples of unsupervised algorithms are the K-means, and C-means.[11]

Semi-supervised machine learning is where we only have some output data say (Y) labeled in have a large amount of input data say (X). It has some features of both unsupervised learning and supervised learning.[10] It has both the characteristics of supervised and unsupervised learning approaches.[13] Here, the data to be trained unlike the unsupervised are labeled normal only, which is termed "one-class" classification.

In the semi-supervised learning we label only portions of the data when acquiring the data or by the help of human experts.[8] It has more real world applications than the supervised learning since they don't require labeling the dataset. In detecting faults in space craft, an accident scenario would be detected as an anomaly, which hard to model.[14] The approach used here is to put up a model that corresponds to standard or existing behaviour, then use the built model for anomaly detection while testing. Semi-supervised learning isn't common because of the difficulty in getting dataset that considers all possible anomalous behaviours. Examples are one-class SVMs and auto-encoders algorithms.

Supervised machine learning has an input variable say (X) mapped to a corresponding output variable say (Y) and then use algorithm to learn the function to map the input to the output (Y=f(X)). Examples of this type of learning are the regression and classification problems.[10]

Here, we label the dataset stating which is normal and which is anomalous, and then construct the predictive model.[11] Mapping out the input features to an output called class is the major task. However, the supervised machine is virtually not so pertinent or applicable due to presumptions that anomalies already are well-known and labelled correctly.

The supervised approach is applicable mostly in problems involving regression or classification. Examples of supervised algorithms are the supervised neural networks, support vector machines (SVM), Bayesian, and K-nearest neighbours' decision tree.

Hybrid learning refers to architectures that make use of generative (unsupervised) as well as discriminative (supervised) components. The combination of different architectures can be used to design a hybrid deep neural network. They are used for action recognition of humans using action bank features and are expected to produce much better results.[15]

A study investigated the prevalence of risk factors for DM in Nigeria.[16] In conducting this research, a total of 23 studies (N=14,650 persons) were considered. In estimating the pooled prevalence of DM, a random-effects model was implemented, and a subgroup-specific DM prevalence was used to account for inter-study and intra-study heterogeneity. The results show that the frequency of DM in Nigeria has been on the increase in all affected regions of the country, with the south-south region having the highest degree in the geopolitical zones. Urbanization, physical inactivity, aging and unhealthy diet were identified as key risk factors for DM amongst Nigerians.[16]

A ML model has been developed to predict DM at an early stage in individuals residing in Northern Nigeria.[17] The study considered a sample of 255 persons which consist of two parts, namely; with 105 diabetic samples and 150 non-diabetic samples. Nine physical DM attributes: age, sex, number of pregnancies, glucose level, blood pressure level, body mass index, height, weight and how regularly they exercise were examined and three supervised learning algorithms of K nearest neighbors (KNN), decision trees, and artificial neural networks (ANN) were used to predict DM from a locally collected dataset in Kaduna State, Nigeria. Their result indicated that ANN produced the highest accuracy, at 97.40%. However not all the DM causing factors such as alcohol, smoking, lack of sleep, excessive sleep, and hereditary that are found to be the predominant life style of those in other regions, especially south southern Nigeria were used in the study.[17]

A stacking ensemble learning approach has been proposed to predict the rate of diabetes occurrence in Maiduguri.[18] Dataset which consisted of 9 distinct features and 1030 individual cases were gathered from individuals aged 17 and above residing in Maiduguri, Borno State, Nigeria, and the surrounding areas, encompassing both males and females. The features considered included pregnancy, glucose levels, blood pressure, skin thickness, insulin levels, body mass index (BMI), diabetes pedigree function, and age. The proposed models included adaptive boosting regression (Adaboost), gradient boosting regression (GBOOST), random forest regression (RFR), ordinary least square regression (OLS), least absolute shrinkage selection operator regression (LASSO), and ridge regression (RIDGE). Performance metrics considered in this study are root mean square (RMSE), mean absolute error (MAE), and mean square error (MSE), which were used to evaluate the performance of the ML and the proposed stacking ensemble learning (SEL) technique. Experimental results revealed that SEL is a better predictor compared to other machine learning approaches considered with an RMSE of 0.0493; a MSE of 0.0024; and a MAE of 0.0349.[18]

A study developed and compared the performance of logistic regression (LR), ANN, and decision tree (DT) models for predicting diabetes mellitus in a Nigerian population, utilizing socio-demographic factors (age, sex, and ethnicity), lifestyle, and physical activities.[19] Accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUROC) were used as performance evaluation metrics. Analysis and model development were performed in R version 3.6.1. The mean age of the participants was 50.52±16.14 years. Experimental result showed that decision trees were the best-performing classifiers with 99.05% and AUC value, 99.2% respectively.[19]

A study developed a predictive supervised machine learning models based on logistic regression, support vector machine, K-nearest neighbor, random forest (RF), naive Bayes and gradient booting algorithms (GNA) for diagnostic dataset of type 2 diabetes mellitus with 383 instances and nine attributes from the Murtala Mohammed Specialist Hospital, Kano.[20] The random forest predictive learning-based model appeared to be one of the best developed models with 88.76% in terms of accuracy; however, in terms of receiver operating characteristic curve, random forest and gradient booting predictive learning-based models were found to be the best predictive learning models with 86.28% predictive ability, respectively.[20]

Another study proposed a predictive data mining model for diabetes mellitus prediction based on the dataset collected from the seven northwestern states of Nigeria.[21] The data were collected from both primary and secondary sources through questionnaires and verbal interviews from patients with diabetic mellitus and other chronic diseases. The dataset comprises of 281 instances with 8 attributes. R programming software (version 5.3.1) was used for implementation. Binomial logistic regression (BLR), classification, confusion matrix and correlation coefficient were to develop the predictive model. The data were partitioned into training and testing sets. Training data were used in building the model while testing data were used to validate the model. The algorithm for the best-fitted model converges with null deviance: 281.951, residual deviance: 16.476 and AIC: 30.476. The significance variables are age, glucose level, diastolic blood pressure level and symptoms associated with kidney diseases with 0.025, 0.01, 0.05 and 0.025 p values, respectively. The predicted model accounted for the accuracy of ~97.1%. The correlation analysis results revealed that diabetic patients are more likely to be hypertensive than patients with other chronic diseases considered in the research.[21]

METHODS

We conducted a systematic review in line with the "preferred reporting items for systematic reviews and meta-analyses" (PRISMA) checklist.[22]

Search strategy

Forty journals from scholarly databases like Google Scholar, PubMed, Science Direct, IEEE Xplore, Springer and Research-Gate were searched for articles. Other health-related databases. Hand-searching of references within included articles was conducted to shortlist other potential articles. Our search strategy utilized a combination of subject terms related to "machine learning," "prediction," and "diabetes complications"

Eligibility criteria

Full-text, English language articles on ML models in an adult (age ≥18 years old) type 2 diabetes population were included. Areas of interest or outcome were prediction and classification of diabetes in Nigeria. Case reports, case series, irrelevant reviews, and meta-analyses were excluded. We also excluded diagnostic ML models and prognostic ML models that predicted diabetes complications. Logistic regression, penalized regression, and generalized additive models were not considered as ML methods in our review and were excluded.[23]

In this review, prognostic ML models refer to models that predict the probability of the future occurrence of the disease in an individual, while diagnostic models predict the disease status of an individual.

Study selection

Two independent reviewers (U.O. and U.J.O.) reviewed the abstracts of retrieved articles and assessed the full text of relevant studies for eligibility. Disagreements during the selection process were discussed to reach a consensus. A third and fourth independent reviewer (O.L.C, and P.B.) was consulted for arbitration of unresolved disagreements.

Data extraction/presentation

Data were extracted using a standardized form comprising items from CHARMS and "transparent reporting of a multivariable prediction model for individual prognosis or diagnosis" (TRIPOD) guidelines.

CHARMS was designed to guide the systematic review of prediction modeling studies and provides a list of relevant items to extract from studies, while TRIPOD comprises a checklist of 22 items developed to guide the reporting of prediction models. Corresponding authors of included studies were contacted for additional details when required.

The information extracted was in two phases. The first phases included the authors name and date of publication, study purpose, diabetes mellitus attributes considered and the study area. The second phase included machine learning method used for prediction, evaluation metrics and findings. This information was presented in tables.

Assessment of bias

The quality of included studies was assessed by four independent reviewers (U.O., U.J.O., A.E.O and O.N.M.) for risk of bias using the "prediction model risk of bias assessment tool" (PROBAST).[24] All disagreements were resolved through discussions with a fifth independent reviewer (P.B.).

RESULTS

Table 1 shows the classification of the related works reviewed itemizing the research purpose, diabetes attributes considered and study area showing the state and region the population of study was drawn from.

Table 2 shows the classification of the related works based the algorithm used, research findings and gaps.

Table 1: Tabular view showing their research purpose, attributes considered and study area.

Authors/dates	Research purpose	DM attributes considered	Study area	
			State	Region
Uloko et al[16]	Examined the prevalence of risk factors for DM in Nigeria using random-effects model	Urbanization, physical inactivity, aging and unhealthy diet	All	All
Evwiekpaefe et al[17]	Developed a machine learning model that predicts DM in individuals at an early stage	Pregnancies, glucose level, blood pressure level, body mass index, height, weight, and regular exercise	Kaduna	North West
Dada et al[18]	Proposed a stacking ensemble learning approach to predict the rate of occurrence of diabetes cases	Pregnancy, glucose levels, blood pressure, skin thickness, insulin	Borno	North West

Continued.

Authors/dates	Research purpose	DM attributes considered	Study area	
			State	Region
		levels, body mass index, diabetes pedigree function, and age		
Odukoya et al[19]	Developed and compared some supervised ML to predict DM	Socio-demographic (age, sex, and ethnicity), lifestyle and physical activities	Lagos	South West
Muhammad et al[20]	Developed a predictive supervised machine learning models for diagnostic dataset of type 2 DM	Type 2 diabetes	Kano	North West
Uba et al[21]	Proposed a predictive data mining model for diabetes mellitus detection	Age, glucose level, diastolic blood pressure level and kidney disease symptoms	Sokoto, Kebbi, Zamfara, Katsina, Kaduna, Kano and Jigawa	North West

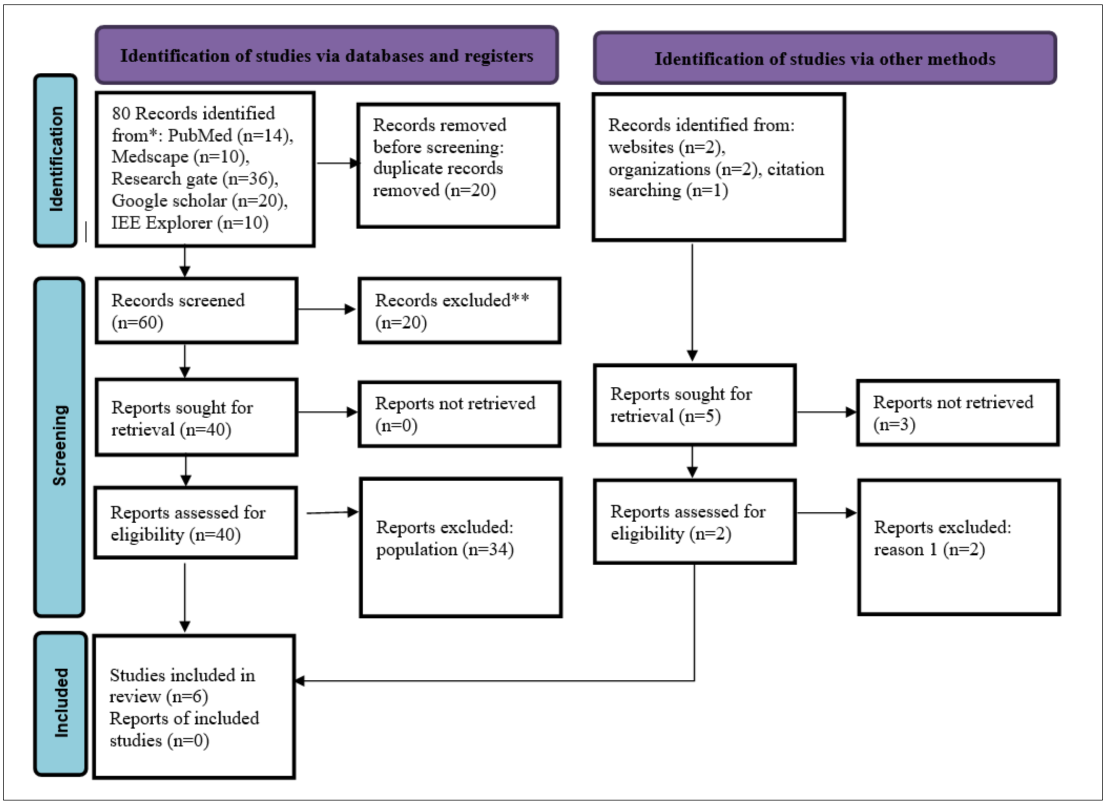

Figure 1: PRISMA 2020 flow diagram for the study which included searches of databases, registers and other sources.[25]

Table 2: Tabular view showing the algorithm used, evaluation metrics' and research findings.

Authors/dates	Algorithm/technique used	Evaluation metrics	Findings
Uloko et al[16]	Random-effects model		Results show that the frequency of DM in Nigeria has been on the increase in all affected regions of the country, with the south-south region having the highest degree in the geopolitical zones
Evwiekpaefe et al[17]	K nearest Neighbors (KNN), Decision Trees, and Artificial Neural Networks (ANN)	Accuracy, precision, recall, F1-ratio	ANN produced the highest accuracy, at 97.40% in terms of prediction

Continued.

Authors/dates	Algorithm/technique used	Evaluation metrics	Findings
Dada et al[18]	Stacking ensemble learning approach	Root mean square (RMSE), mean absolute error (MAE), mean square error (MSE)	SEL is a better predictor compared to other machine learning approaches considered with an RMSE of 0.0493; a MSE of 0.0024; and a MAE of 0.0349.
Odukoya et al[19]	Logistic regression (LR), artificial neural network (ANN), and decision tree (DT)	Accuracy, sensitivity, specificity, PPV, NPV, AUROC	Experimental result showed that decision trees were the best-performing classifiers with 99.05% and AUC value, 99.2% respectively.
Muhammad et al[20]	Logistic regression, support vector machine, K-nearest neighbor, random forest (RF), naive Bayes and Gradient booting algorithms (GNA)	Accuracy	The random forest predictive learning-based model appeared to be one of the best developed models with 88.76% in terms of accuracy
Uba et al[21]	Binomial logistic regression (BLR), classification, confusion matrix and correlation coefficient	Accuracy	The predicted model accounted for the accuracy of ~97.1%. The correlation analysis results revealed that diabetic patients are more likely to be hypertensive than patients with other chronic diseases.

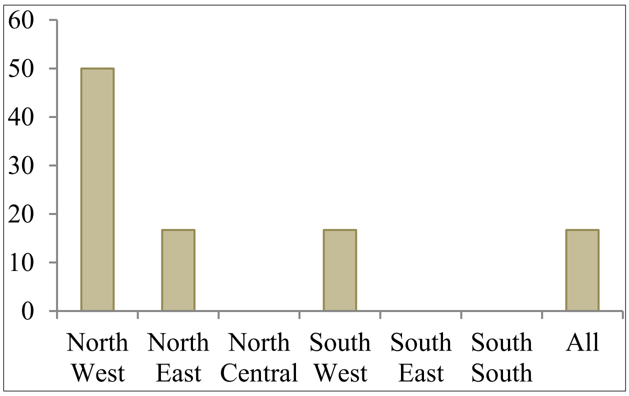

Figure 2: Chart showing the percentage distribution of region the population of study.

DISCUSSION

Results from Table 1 shows that much work has not been carried on the prediction and classification of diabetes using machine learning in Nigeria. Despite the few works that have been done, Figure 2 shows that mostly the Northern population (66.7%) has been investigated, and they employed various research methodologies, including random-effects models, machine learning algorithms, and stacking ensemble learning approaches. The review also showed geographic diversity wherein studies were conducted in different regions of Nigeria, including the North West (Kaduna, Kano), North East (Borno), South West (Lagos), and All (covering the entire country), providing insights into regional variations in DM prevalence and risk factors.

As regards prevalence, it was gleaned that there is an increasing prevalence of DM and the growing number of studies on DM in Nigeria suggests a growing concern about the disease's prevalence and impact on the population, a study carried out a review using ML on the prevalence of diabetes on the six geo-political zones in Nigeria and concluded that the south southern region as highest degree of prevalence in the geopolitical zones no work has been considered on the prediction and classification of ML on the south southern population.[16] Collecting data in these regions may require careful consideration of ethical guidelines and obtaining informed consent. Focus was also placed on identifying the prevalence of DM and its associated risk factors, including urbanization, physical inactivity, unhealthy diet, various socio-demographic factors, and lifestyle behaviors.[16,19]

The predominant risk factors consistently examined in the literature include glucose level, blood pressure, age, BMI, and exercise [Ma1]. Other factors explored, as detailed in Table 1, encompass symptoms of kidney disease, insulin level, pregnancy, sex, ethnicity, skin thickness, and urbanization. However, key risk factors prevalent in the Nigerian population, particularly in the southern region, such as alcohol consumption, smoking, dietary habits, and irregular sleep patterns, have been largely overlooked.[26] Given that the South-South region exhibits the highest prevalence of diabetes in Nigeria, this disparity in examined factors may be significant.[16] The elevated prevalence in the southern and eastern regions could be attributable to environmental factors, dietary norms, and lifestyle choices, including smoking, alcohol consumption, and sleeping habits, which are more common in these areas.

While numerous studies globally and in other African nations have leveraged ML for diabetes prediction, a notable gap exists concerning the southern region of Nigeria. Predominantly, supervised learning algorithms are commonly employed, a trend likely driven by the structured nature of available health data. The continuous monitoring and staging requirements for effective diabetes management often necessitate classification tasks, making algorithms like SVM, KNN, logistic regression, random forest, ANN, Bayesian methods, and decision trees common choices in the Nigerian context.[17,19,21]

Among these algorithms applied to Nigerian datasets, ANN have demonstrated superior predictive accuracy, reaching 97.40%. Decision trees, on the other hand, have shown exceptional performance in classifying diabetes into different stages (positive, negative, early, middle, advanced) with an accuracy of 99.05% and an area under the curve (AUC) of 99.2%.[17,19] Furthermore, logistic regression stands out as the most frequently utilized algorithm for analyzing diabetes-related datasets in Nigeria, particularly in examining risk factors, achieving an accuracy of 97.1%.[19,20]

However, the application of ML in developing countries like Nigeria faces significant challenges. These include limitations in digital infrastructure, inconsistencies in data quality and availability, a shortage of skilled professionals in data science and AI, and regulatory as well as ethical considerations regarding the use of sensitive health data.[27] In Comparison with studies in other African countries like Zimbabwe and East Africa, and globally, reveals a similar reliance on supervised learning but also highlights the context-specific nature of effective models.[28] Addressing the unique challenges within Nigeria is crucial for developing robust and regionally relevant ML models for diabetes prediction and management.

In conclusion, while diabetes is becoming a major health issue in Nigeria, with a rising number of cases, leading to serious complications such as blindness, kidney failure, heart problems, and limb amputations. Over 1.5 million people worldwide have died from diabetes-related complications. Early detection and treatment are essential to prevent these problem, though machine learning has been extensively applied for diabetes prediction globally and in some African countries, research focusing specifically on the southern region of Nigeria remains scarce, despite the area's high prevalence. Supervised learning algorithms like ANN and decision trees have demonstrated promising results for prediction and classification in Nigerian datasets, with logistic regression being a common tool for risk factor analysis. However, the effective implementation of ML in developing countries like Nigeria is hindered by infrastructural limitations, data quality issues, a lack of skilled personnel, and ethical considerations, necessitating targeted efforts to address these challenges for the development of contextually relevant and impactful solutions for diabetes management in the region

Based on the finding of the review, a hybrid user-friendly decision support system for diabetes management, accessible in remote areas via mobile phones was recommended. Further research is needed to explore regional variations and genetic factors in diabetes susceptibility. This will enable personalized risk assessment, prevention strategies, contributing to effective diabetes management in Nigeria.

CONCLUSION

This review highlights a critical gap in Nigerian machine learning diabetes research: the lack of focus on the high-prevalence South-South region. While identifying key risk factors and confirming the high accuracy of ANNs (97.40%) and decision trees (99.05%) for prediction, it stresses that vital regional factors like diet and smoking are overlooked. The study advocates for targeted research, mobile-accessible decision support systems, and addressing data/infrastructure challenges for effective, context-specific diabetes management.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. American Diabetes Association. What is diabetes? 2023. Available at: https://diabetes.org/. Accessed on 12 April 2025.
2. Adebayo EF, Owolabi OD, Fagbenle OO. Prevalence and risk factors for diabetes mellitus in a semi-urban Nigerian community: A cross-sectional study. Int J Environ Res Public Health. 2020;17(13):4882.
3. Pradhan N, Rani G, Dhaka VS, Poonia RC. Diabetes prediction using artificial neural network. In: Deep Learning Techniques for Biomedical and Health Informatics. Elsevier. 2020;327-39.
4. Liu J, Tang ZH, Zeng F, Li Z, Zhou L. Artificial neural network models for prediction of cardiovascular autonomic dysfunction in general Chinese population. BMC Med Inform Decis Mak. 2013;13(1):80.
5. Standl E, Khunti K, Hansen TB, Schnell O. The global epidemics of diabetes in the 21st century: Current situation and perspectives. Eur J Prev Cardiol. 2019;26(2):7-14.
6. World Health Organization. Diabetes. 2021. Available at: https://www.who.int/health-topics/diabetes. Accessed on 12 April 2025.
7. Tan KR, Seng JJB, Kwan YH, Chen YJ, Zainudin SB, Loh DHF, et al. Evaluation of machine learning methods developed for prediction of diabetes complications: A systematic review. J Diabetes Sci Technol. 2023;17(2):474-89.
8. Buczak AL, Guven E. A survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Commun Surv Tutor. 2016;18(2):1153-76.
9. Witten IH, Frank E. Data mining: Practical machine learning tools and techniques. Morgan Kaufmann. 2005.
10. Brownlee J. Supervised and unsupervised machine learning algorithms. 2016. Available at: https://machinelearningmastery.com/supervised-

and-unsupervised-machine-learning-algorithms/. Accessed on 12 April 2025.
11. Omar S, Ngadi A, Jebur HH. Machine Learning Techniques for Anomaly Detection. Int J Comput Appl. 2013;69(9):33-41.
12. Guobing Z, Cuixia Z, Shanshan S. A Mixed Unsupervised Clustering-based Intrusion Detection Model. In: 2009 Third International Conference on Genetic and Evolutionary Computing. IEEE. 2009;402-5.
13. Van Engelen JE, Hoos HH. A survey on semi-supervised learning. Mach Learn. 2020;109:373-440.
14. Fujimaki R, Yairi T, Machida K. An approach to spacecraft anomaly detection problem using kernel feature space. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. ACM. 2005;401-10.
15. Li KC, Wong BTM, Kwan R, Chan HT, Wu MMF, Cheung SKS. Evaluation of hybrid learning and teaching practices: The perspective of academics. Sustainability. 2023;15(8):6780.
16. Uloko AE, Musa BM, Ramalan MA, Gezawa ID, Puepet FH, Uloko AT, et al. Prevalence and risk factors for diabetes mellitus in Nigeria: A systematic review and meta-analysis. Diabetes Ther. 2018;9(3):1307-16.
17. Evwiekpaefe AE, Abdulkadir N. A predictive model for diabetes mellitus using machine learning techniques (A study in Nigeria). Afr J Inform Syst. 2023;15(1):1-21.
18. Dada EG, Birma AI, Gora AA. Ensemble machine learning algorithm for diabetes prediction in Maiduguri, Borno State. Mikailalsys J Math Stat. 2024;2(2):46-73.
19. Odukoya O, Nwaneri S, Odeniyi I, Akodu B, Oluwole E, Olorunfemi G, et al. Development and comparison of three data models for predicting diabetes mellitus using risk factors in a Nigerian population. Healthc Inform Res. 2022;28(1):58-67.
20. Muhammad LJ, Ebrahem AA, Sani SU. Predictive supervised machine learning models for diabetes mellitus. SN Comput Sci. 2020;1(4):240.
21. Uba MM, Jiadong R, Sohail MN, Irshad M, Yu K. Data mining process for predicting diabetes mellitus based model about other chronic diseases: a case study of the northwestern part of Nigeria. Healthc Technol Lett. 2019;6(4):98-102.
22. Tan KR, Seng JJB, Kwan YH, Chen YJ, Zainudin SB, Loh DHF, et al. Evaluation of machine learning methods developed for prediction of diabetes complications: A systematic review. J Diabetes Sci Technol. 2023;17(2):474-89.
23. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22.
24. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-8.
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. J Clin Epidemiol. 2021;134:103-12.
26. Selvin E, Parrinello CM. Age-related differences in glycaemic control in diabetes. Diabetologia. 2013;56(12):2549-51.
27. Oduoye MO, Fatima E, Muzammil MA, Dave T, Irfan H, Fariha FNU, et al. Impacts of the advancement in artificial intelligence on laboratory medicine in low- and middle-income countries: Challenges and recommendations-A literature review. Health Sci Rep. 2024;7(1):e1794.
28. Ngusie HS, Tesfa GA, Taddese AA, Enyew EB, Alene TD, Abebe GK, et al. Predicting place of delivery choice among childbearing women in East Africa: a comparative analysis of advanced machine learning techniques. Front Public Health. 2024;12:1439320.

Cite this article as: Ukoba O, Ukoba OJ, Ochei LC, Peter-Kio BO, Anuku EO, Obi-Ntumeonuwa M. A systematic review of machine learning methods for diabetes mellitus prediction and classification in Nigeria. Int J Community Med Public Health 2025;12:2828-35.