Original Article

Prediction of Diabetes in Middle-Aged Adults: A Machine Learning Approach

Gideon Addo[1,2], Bismark Amponsah Yeboah[1], Michael Obuobi[3], Raphael Doh-Nani[1,2], Seidu Mohammed[1,2], David Kojo Amakye[3]

[1]Department of Statistics and Actuarial Science, College of Science, Kwame Nkrumah University of Science and Technology, [2]Department of Pathology, School of Medicine and Dentistry, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, [3]Department of Mathematical Sciences, College of Science, University of Texas at El Paso, El Paso, Texas, USA

Abstract

Background: Diabetes is a serious health concern requiring effective diagnostic strategies, particularly since its symptoms overlap with those of other conditions. Despite extensive research on early diabetes detection across various age groups, middle-aged adults have been relatively underexplored. This study focuses on this demographic to examine symptom-diabetes associations, examine the influence of symptoms in diabetes prediction, and determine an optimal machine learning (ML) model for diabetes prediction. **Materials and Methods:** This study utilized data from a previous cohort study conducted in Bangladesh. The original dataset included demographic and symptom-related information from 520 patients visiting the ABC Hospital in Bangladesh, India. The participants comprised both diabetic and non-diabetic individuals showing diabetes-like symptoms. For our study, data from 296 middle-aged adults (aged 40–60 years) were extracted. Chi-square tests assessed diabetes-symptom associations, and the Boruta algorithm examined feature influence. Seven ML classification models were evaluated for predictive accuracy. **Results:** Results showed that 60% of the 296 participants were diabetic. Symptoms like polyuria, polydipsia, weakness, sudden weight loss, partial paresis, polyphagia, and visual blurring were significantly associated with diabetes. All demographic and symptom-related features were influential in diabetes prediction, with polyuria, polydipsia, gender, alopecia, and irritability emerging as the most influential. Among the ML models tested, the random forest model exhibited the highest sensitivity (98.59%) and outperformed others in accuracy (96.58%) and area under the curve score (96.00%), making it the most efficient model for predicting diabetes in middle-aged adults. **Conclusion:** Diabetes associated symptoms provide valuable diagnostic opportunities for early diabetes detection in middle-aged adults. Future research should explore genetic, lifestyle, and environmental factors to improve diagnostic accuracy.

Keywords: Diabetes prediction, diabetes symptoms, machine learning models, middle-aged adults, predictive performance metrics

Key Messages: Our study reveals that diabetes-associated symptoms provide valuable diagnostic opportunities, enabling accurate diabetes predictions in middle-aged adults.

Introduction

Diabetes stands as a persistent medical condition marked by heightened blood glucose levels and disruptions in the metabolism of fats and proteins.[1] This chronic ailment encompasses a cluster of symptoms linked to hyperglycemia, signifying elevated sugar levels in the bloodstream.[2] The complex process by which the body converts food into energy is altered in diabetes. Typically, the body breaks down food into sugar and glucose, releasing them into the bloodstream. Elevated blood sugar levels prompt the pancreas to release insulin, acting as a key to facilitate the entry of sugar into the

Received: 28-Jun-2024, **Revised:** 19-Aug-2024, **Accepted:** 26-Aug-2024, **Published:** 30-Oct-2024

Access this article online

Quick Response Code:

Website: https://journals.lww.com/JODB

DOI: 10.4103/jod.jod_103_24

Address for correspondence: Mr. Gideon Addo, Department of Statistics and Actuarial Science, College of Science, Kwame Nkrumah University of Science and Technology, Private Mail Bag, University Post Office, Kumasi 00233, Ghana
E-mail: addogideon41@gmail.com

How to cite this article: Addo G, Yeboah BA, Obuobi M, Doh-Nani R, Mohammed S, Amakye DK. Prediction of diabetes in middle-aged adults: A machine learning approach. J Diabetol 2024;15:401-8.

body's cells for energy utilization.[3] In diabetes, either insufficient insulin is produced, or the body fails to use insulin effectively, leading to excessive blood sugar lingering in the bloodstream. Over time, this condition can give rise to severe health complications such as heart disease, vision impairment, and kidney disorders.[3] The global prevalence of diabetes has surged, quadrupling since 1980, with an estimated 422 million people worldwide affected, according to the World Health Organization.[4]

The Centers for Disease Control and Prevention identifies three primary types of diabetes: type 1 (T1D), type 2 (T2D), and gestational (GD) diabetes. T1D is a heterogeneous disorder characterized by the destruction of pancreatic beta cells, resulting in absolute insulin deficiency.[5] As an autoimmune disorder, T1D is considered the most common and severe among autoimmune diseases,[6] though it accounts for only 5%–10% of global diabetes cases.[7,8] T1D can manifest at any age, with the highest incidence observed in children.[9] On the other hand, T2D is a progressive disorder marked by deficiencies in insulin secretion and increased insulin resistance, leading to abnormal glucose metabolism and related metabolic disruptions.[10] Representing approximately 90% of all diabetes cases, T2D predominantly affects middle-aged adults.[11] Symptoms of T2D may develop gradually over several years, sometimes going unnoticed for an extended period.[12] Lastly, GD is a less common form of diabetes occurring in pregnant women without preexisting diabetes, leading to elevated blood sugar levels during pregnancy. The outcomes of GD include an elevated risk of maternal cardiovascular disease and T2D.[13] Additionally, infants born to mothers with GD may experience macrosomia (excessive birth weight) and complications during birth.[13] Globally, 14% of all pregnancies are affected by GD.[14]

Irrespective of the diabetes type, neglecting early signs and symptoms can result in severe long-term complications, including cardiovascular issues, kidney problems, vision impairment, and neuropathic conditions. Recognizable early signs that necessitate diabetes diagnosis and investigation include frequent urination (polyuria), excessive thirst (polydipsia), sudden weight loss, excessive hunger (polyphagia), blurred vision, unexpected weakness, dry skin, delayed healing, and frequent infections.[15] Given the overlap of symptoms with other medical conditions, there is a crucial need for highly efficient and user-friendly diagnostic methods to aid healthcare professionals and patients in determining the diabetes status of individuals presenting with similar symptoms.

In light of the escalating global incidence and mortality rates of diabetes,[16] numerous studies have delved into understanding the etiology, diagnosis, and management of the condition. For the effective management of diabetes, early detection is crucial, as the condition can deteriorate over time if not properly addressed or remains undetected. While many diagnostic studies rely on laboratory-based tests, some are outdated [17-19] despite some having high accuracy rates.[19] The recent surge in artificial intelligence has prompted a growing number of studies utilizing machine learning (ML) techniques to detect and screen diabetes across various demographics and populations, including gender,[20-22] income classes,[23] and age groups such as infants,[24] children, and teenagers.[25-27] Surprisingly, few, if any, of these studies explicitly concentrate on employing ML techniques to screen and diagnose diabetes in middle-aged adults. This study aims to fill this gap by focusing on using ML methods to detect diabetes, specifically in the middle-aged adult population.

The primary objectives of this study are to (1) assess associations between symptoms displayed by middle-aged adults and their diabetes status, (2) investigate the relevance and relative influence of some demographic and symptomatic features in predicting diabetes status among middle-aged adults, and (3) identify the most effective ML model for predicting diabetes status in middle-aged adults. The significance of this study lies in its potential to provide valuable insights into the associations between symptoms, demographic factors, and diabetes status in the middle-aged population.

Materials and Methods

Study design and data source

This research utilized the dataset derived from a previous study conducted by Islam *et al.*[28] in India, where they aimed at constructing a predictive tool for early-stage diabetes risk predictions. The dataset, publicly available on Mendeley Data,[29] includes information on patient demographics and diabetes-related symptoms obtained from 520 participants through a questionnaire. These participants were patients visiting the Sylhet Diabetes Hospital in Bangladesh and included both diabetic and non-diabetic individuals showing signs similar to that of diabetes. The dataset contains comprehensive information on patient symptoms, demographics, and diabetes status, making it suitable for addressing the objectives of this current investigation.

For our study, we specifically extracted data for 296 middle-aged adults within the original dataset, where the inclusion criterion was for individuals aged between 40 and 60 years. The exclusion criterion was for all other patients who fell outside this age bracket.

Features and feature selection

Within the dataset employed for this study, 17 variables were identified, with 16 serving as features and one as the outcome variable. The outcome variable contained information about a patient's diabetes status, while the

features included details on patient demographics and symptoms linked to diabetes, as outlined in Table 1. All features, excluding the "age" variable, were nominal with two classes. Notably, our dataset contained no missing values.

To assess the influence of features on predicting diabetes status, we employed the Boruta algorithm, accessible through the Boruta package in R,[30] for feature selection. The Boruta algorithm, recognized for its all-relevant approach, utilizes a wrapper algorithm to identify influential features. Leveraging the random forest (RF) algorithm, it conducts a comprehensive search for relevant features by comparing original variable importance with estimates derived from permuted copies, progressively eliminating irrelevant features. This algorithm is adaptable, accommodating any classification method outputting variable importance.

Table 1: Overview of the features used in this research to study diabetes among middle-aged adults

Feature	Description
Age	Age of patient
Sex	Gender of patient
Polyuria	Signs of frequent urination
Polydipsia	Signs of extreme thirstiness
Polyphagia	Signs of extreme hunger
Sudden weight loss	Unexplained weight loss
Genital thrush	Recent genital yeast infection
Visual blurring	Blurred vision or poor eyesight
Itching	Frequent itching of the body
Irritability	Feeling agitated or frustrated
Delayed healing	Poor healing of wounds
Partial paresis	Partial loss of voluntary movement in the limbs
Muscle stiffness	Stiff muscles
Alopecia	Sudden unexplained hair loss
Obesity	Obese patient
Diabetes status	Diabetes status of the patient

For a visual assessment of feature influence and relevance, we utilized a Boruta algorithm-generated plot, illustrated in Figure 1. In this plot, the vertical axis represents feature importance, while the horizontal axis represents features. The color of each boxplot indicates the relevance of the feature in predicting the outcome variable. Blue boxplots correspond to the minimal, average, and maximum Z score of a shadow feature, while red, green, and yellow boxplots represent Z scores of rejected, confirmed, and tentative features, respectively. Confirmed features are those recognized as significant predictors of the outcome variable. Rejected features, on the other hand, are deemed non-important in predicting the outcome variable. Tentative features are predictors for which the Boruta algorithm could not conclusively determine their importance.

Statistical analysis

To prepare our data for analysis, we conducted data preprocessing. Except for the "age" variable, all other variables were nominal with two classes. Utilizing one-hot encoding, we encoded all nominal variables, representing the presence of a category with "1" and its absence with "0." Additionally, we applied the min-max scaling technique to standardize the "age" variable's scale, bringing its values within the range of 0 to 1.

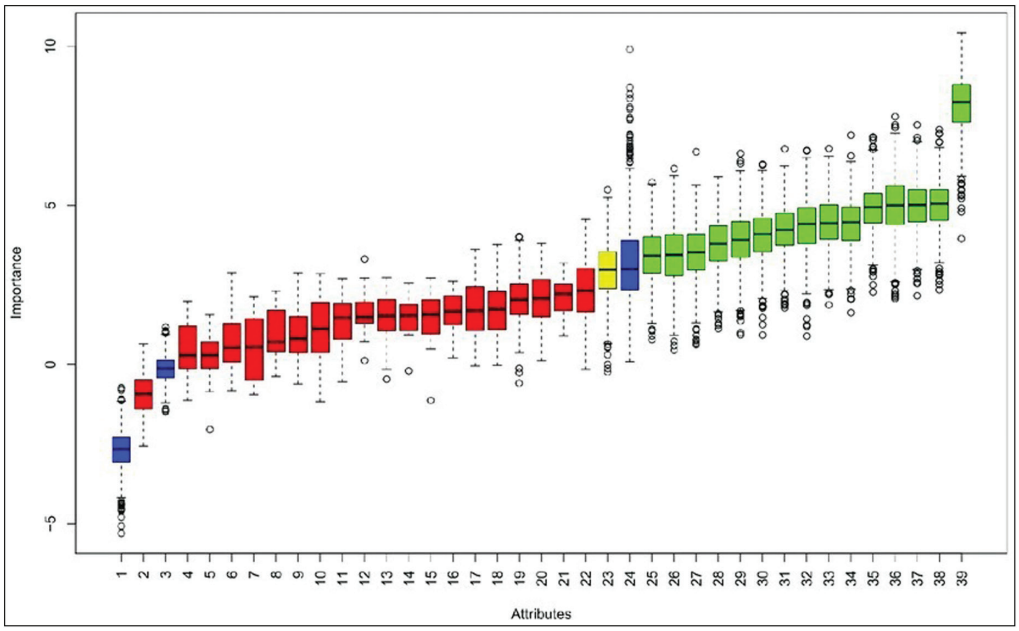

Figure 1: Illustration of a sample Boruta algorithm-generated feature importance plot

In this study, descriptive statistics were employed to characterize the primary outcome variable and all other predictor variables. Frequencies and percentages were used to describe the nominal variables. To examine associations between symptoms exhibited by middle-aged adults and their diabetes status, we employed the Chi-square test of association. This statistical method evaluates the dependence or independence between two nominal variables, determining whether a significant association exists based on observed and expected frequencies within a contingency table.

Furthermore, to determine the most efficient ML algorithm for predicting diabetes within the middle-aged adult population, we assessed the performance of seven distinct algorithms. These algorithms include K-nearest neighbor (KNN), naïve Bayes classifier, support vector machines (SVM) with linear, polynomial, and radial basis function kernels, RF classifier, and logistic regression (LR). We employed 10-fold cross-validation to select the best-performing model for each type of ML algorithm to ensure robust model evaluation and minimize the risk of overfitting. In evaluating the performance of the ML models, various performance metrics were utilized, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and receiver operating characteristic (ROC) curve analysis with their associated area under the curve (AUC) values. In this study, the selection of the optimal ML model for predicting diabetes status among middle-aged adults was based on the one demonstrating both the highest accuracy and AUC scores.

The chosen level of significance throughout the study was set at 0.05. All statistical analyses were carried out using the R programming language (version 4. 3. 2, R Core Team, Vienna, Austria).

Ethical Approval

We did not seek ethical approval from the Committee on Human Research and Publication Ethics (CHRPE) at the School of Medicine and Dentistry, Kwame Nkrumah University of Science and Technology, because the data utilized in this study was secondary and publicly available under a Creative Commons Attribution 4.0 International license. The dataset is anonymized, containing no personally identifiable information, and we were not involved in the original data collection process. In compliance with the principles outlined in the Declaration of Helsinki, we ensured that all ethical standards regarding research on human data were maintained throughout the study.

Results

Assessing the associations between symptoms displayed by middle-aged adults and their diabetes status

Table 2 displays symptoms in middle-aged patients, both with and without diabetes, along with the *P* value from Chi-square tests assessing associations with diabetes status. Among the 296 participants, 179 (60%) were diabetic. Symptoms like polyuria, polydipsia, weakness, sudden weight loss, partial paresis, polyphagia, and visual blurring are more prevalent in diabetic middle-aged patients and rare in their non-diabetic counterparts. This emphasizes the significance of these symptoms as potential indicators of diabetes in this age group.

The *P* value from the Chi-square tests highlights strong associations between diabetes status and symptoms like polyuria, polydipsia, weakness, sudden weight loss, partial paresis, polyphagia, and visual blurring in middle-aged individuals. These results emphasize the importance of these symptoms as strong indicators of diabetes in this age group. In contrast, symptoms such as obesity, delayed

Table 2: Summary results of the Chi-square tests of association between diabetes status and symptoms among middle-aged adults

Symptom	Diabetes (n = 179)		Diabetes free (n = 117)		*P* value
	Yes (%)	No (%)	Yes (%)	No (%)	
Polyuria	136 (76)	43 (24)	4 (3)	113 (97)	0.000
Polydipsia	134 (75)	45 (25)	8 (7)	109 (93)	0.000
Sudden weight loss	125 (70)	54 (30)	23 (20)	94 (80)	0.000
Weakness	132 (74)	47 (26)	64 (55)	53 (45)	0.001
Polyphagia	107 (60)	72 (40)	29 (25)	88 (75)	0.000
Genital thrush	43 (24)	136 (76)	30 (26)	87 (74)	0.859
Visual blurring	103 (58)	76 (42)	36 (31)	81 (69)	0.000
Itching	84 (47)	95 (53)	74 (63)	43 (37)	0.008
Irritability	53 (30)	126 (70)	7 (6)	110 (94)	0.000
Delayed healing	85 (47)	94 (53)	62 (53)	55 (47)	0.420
Partial paresis	114 (64)	65 (36)	22 (29)	95 (71)	0.000
Muscle stiffness	75 (42)	104 (58)	40 (34)	77 (66)	0.000
Alopecia	34 (19)	145 (81)	77 (66)	40 (34)	0.000
Obesity	39 (22)	140 (78)	23 (20)	94 (80)	0.769

healing, and genital thrush showed no association with diabetes status in middle-aged individuals, suggesting that these conditions may not be as specific to diabetes in this age group.

Investigating the relevance and relative influence of the demographic and symptomatic features in the prediction of diabetes status

Figure 2 visually presents the relevance and relative importance of each variable employed in our study to predict diabetes status in middle-aged adults. The plot reveals clinical and statistical implications that can enhance the understanding and application of these symptoms in the diagnosis of diabetes in this specific population.

From the plot, we observe that certain predictors, namely polyuria, polydipsia, gender, alopecia, irritability, and sudden weight loss, are highly influential features in descending order of importance when it comes to predicting the diabetes status of middle-aged patients. Conversely, features such as genital thrush, itching, obesity, and muscle stiffness are shown to be less important when predicting diabetes status in middle-aged patients. This finding suggests that while these features still play a role in diabetes diagnosis, their relative importance is lower. The green-colored boxplots associated with all features signify that the Boruta algorithm identified each variable as an important contributor to predicting diabetes status in middle-aged patients.

Identifying the most effective ML model for predicting diabetes in middle-aged adults

Table 3 presents a comprehensive evaluation of the performance of the seven ML models used in predicting diabetes in our study, considering their sensitivity, specificity, PPV, NPV, and accuracy scores. From the table, the RF model demonstrated the highest sensitivity at 98.59%, outperforming other models, while KNN

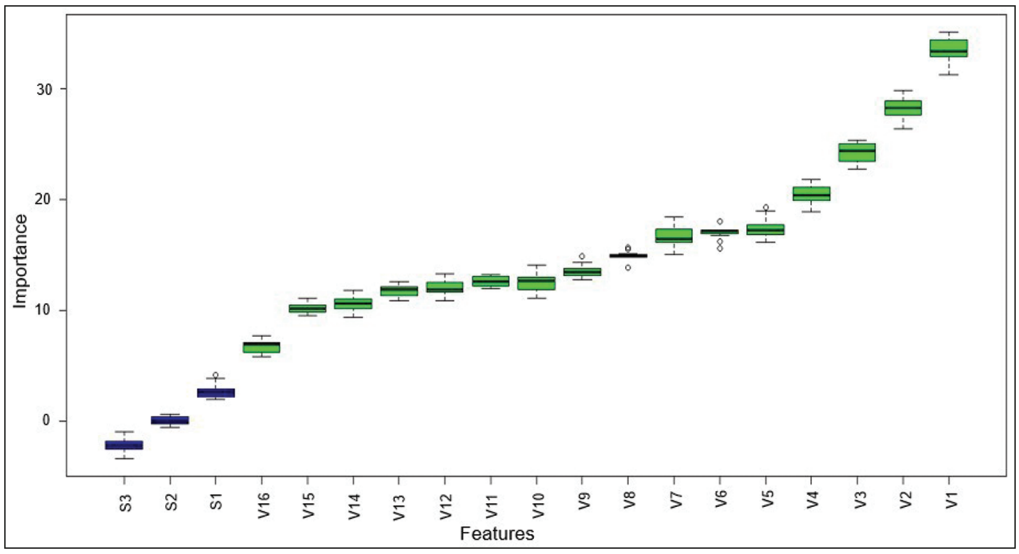

Figure 2: Relative importance of features in predicting diabetes status among middle-aged adults. V1 = polyuria, V2 = polydipsia, V3 = gender, V4 = alopecia, V5 = irritability, V6 = sudden weight loss, V7 = age, V8 = partial paresis, V9 = polyphagia, V10 = delayed healing, V11 = visual blurring, V12 = muscle stiffness, V13 = weakness, V14 = itching, V15 = obesity, V16 = genital thrush, S1 = ShadowMax, S2 = ShadowMean, S3 = ShadowMin

Table 3: Performance metrics associated with each machine learning model in the prediction of diabetes among middle-aged adults

Model	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
KNN (k = 5)	85.92	97.83	98.39	81.82	90.60
NB	97.18	78.26	87.34	94.74	89.74
SVM (linear)	91.55	95.65	97.01	88.00	93.16
SVM (polynomial)	92.96	95.65	97.06	89.80	94.02
SVM (RBF)	91.55	93.48	95.59	87.76	92.31
RF	98.59	93.48	95.89	97.73	96.58
LR	92.96	95.65	97.06	89.80	94.02

KNN = K-nearest neighbor, NB = naïve Bayes, SVM (linear) = support vector machines with linear kernel, SVM (polynomial) = support vector machines with polynomial kernel, SVM (RBF) = support vector machines with radial basis function kernel, RF = random forest, LR = logistic regression, PPV = positive predictive value, NPV = negative predictive value

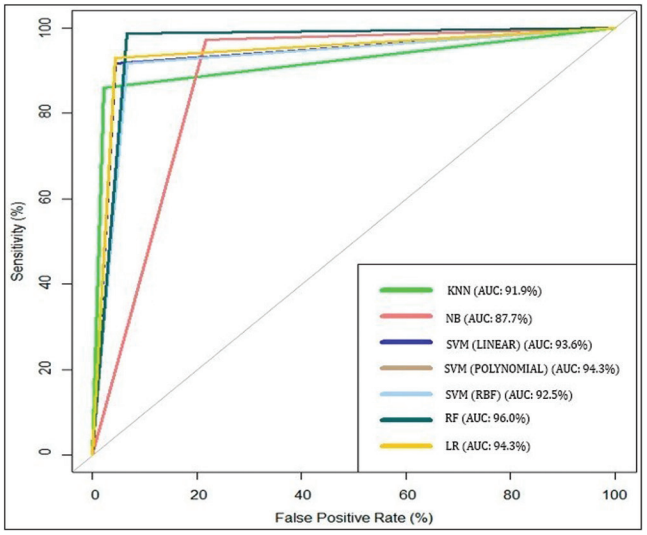

Figure 3: Receiver operating characteristic curves with their corresponding area under the curve scores of the machine learning models of this study. KNN = K-nearest neighbor, NB = naïve Bayes, SVM (linear) = support vector machines with linear kernel, SVM (polynomial) = support vector machines with polynomial kernel, SVM (RBF) = support vector machines with radial basis function kernel, RF = random forest, LR = logistic regression, AUC = area under the curve

ranked highest in specificity with a score of 97.83%. The KNN model achieved the highest PPV at 98.39%, indicating its reliability in identifying diabetic cases, while the RF model had the highest NPV at 97.73%, indicating its reliability in identifying non-diabetic cases. Overall accuracy was highest in the RF model with a score of 96.58%, emphasizing its effectiveness in predicting both diabetic and non-diabetic cases among middle-aged adults.

ROC and AUC analysis

Figure 3 illustrates the ROC curves and their corresponding AUC values for each model used in predicting diabetes among middle-aged adults in this study. AUC values are crucial indicators of a model's ability to distinguish between positive and negative cases. From the plot, the RF model stood out with an impressive AUC of 96.0%, surpassing other models. While AUC values were closely similar, SVM (polynomial) and LR had overlapping curves with identical AUC values of 94.3%. In summary, the RF model exhibited superior discriminative performance, making it the optimal choice for predicting diabetes in middle-aged adults. The consistently high AUC scores of the models highlight the robust predictive capabilities of ML models in the prediction of diabetes.

Discussion

Diabetes is a significant health concern, and its symptoms can remain subtle or unnoticed for extended periods, particularly in T2D.[16,31] Since key diabetes symptoms may resemble those of other medical conditions, individuals often prioritize addressing those conditions. Consequently, diabetes diagnosis tends to occur in later disease stages when symptoms become more severe. This study explored the associations between diabetes status and specific symptoms in middle-aged adults, including both diabetic and non-diabetic individuals. Our findings indicate that among middle-aged adults, the prevalent symptoms associated with diabetes include polyuria, polydipsia, weakness, sudden weight loss, polyphagia, partial paresis, and visual blurring. Conversely, these symptoms are less common among their non-diabetic counterparts.

Several studies across different populations have reported findings that are consistent with ours.[31-34] Devi *et al.*,[31] in their investigation into the prevalence of T2D among the Meiteis of Manipur in India, identified similar symptoms among T2D patients, including polyuria, polyphagia, polydipsia, fatigue, and weight loss, all of which were significantly associated with diabetes. Similarly, Kumar and Kaplowitz[32] observed a parallel discovery in a different age demographic, noting that children with diabetes commonly exhibited symptoms such as polyuria, polydipsia, weight loss, and fatigue. In another study, Pawar *et al.*[33] found a strong indication of diabetes in patients displaying polydipsia, polyuria, polyphagia, or a combination of these symptoms. The consistency of our study's findings with those of previous studies in different populations suggests that middle-aged adults exhibit symptoms of diabetes similar to those in other demographics, with the typical symptoms being polyuria, polyphagia, polydipsia, visual blurring, and fatigue.

In contrast, our study revealed that symptoms such as obesity, delayed healing, and genital thrush were not associated with the diabetes status of middle-aged adults. This contradicts previous research that established a strong association between T2D and obesity.[35-37] For instance, a 2011 study by Barnes[37] found a significant increase in the risk of diabetes among overweight and obese women, with overweight women facing a 28-fold risk and obese women an astonishing 93-fold risk compared to normal-weight women. The reasons for these discrepancies between our study's findings and existing research are not entirely clear. One plausible explanation is that a majority of our study participants may be suffering from T1D, which is typically not associated with obesity.

Moreover, our study found no association between delayed wound healing and diabetes among middle-aged adults, contradicting previous research.[38,39] Fahey *et al.*[38] concluded in their study that delayed healing in diabetes is linked to altered leucocyte infiltration and wound fluid interleukin-6 levels during the late inflammatory phase of the wound healing process. Our findings suggest that certain symptoms traditionally associated with diabetes may not always be indicative in this specific demographic.

Therefore, further research is warranted to explore the factors influencing these differences and refine our understanding of diabetes manifestation in middle-aged adults.

Additionally, we explored the relevance and relative influence of symptoms and demographic factors in predicting diabetes in middle-aged adults. Our study's findings revealed a hierarchy of influential symptoms, with polyuria, polydipsia, alopecia, and irritability being the most critical indicators of diabetes among middle-aged individuals. Gender also played a significant role in this age group. This hierarchy provides valuable information for healthcare professionals, highlighting the crucial relevance of these symptoms in diagnosing diabetes among middle-aged adults. Notably, polyuria and polydipsia emerged as robust early signals for diabetes in this specific population. Conversely, symptoms such as itching, obesity, and muscle stiffness, although identified as important determinants in predicting diabetes among middle-aged adults, had a relatively low influence in predicting diabetes within this age demographic. Healthcare professionals can use this information to optimize diagnostic assessments, potentially expediting the identification of diabetes in middle-aged adults.

Precise classification of diabetes status holds paramount importance, as it prevents the progression of the condition and facilitates potential reversals, especially in cases of prediabetes. The recent surge in AI has spurred numerous research studies aiming to streamline the identification and diagnosis of diabetes, resulting in the need for models with exceptionally high accuracy measures. Several studies employing ML techniques for diabetes predictions consistently yield classification models with outstanding performance scores.[40-46]

In our study, all the ML models exhibited commendable sensitivity scores, signifying their ability to accurately predict diabetes cases among the middle-aged adult population. Notably, the RF model stood out with the highest sensitivity score of 98.59%. The specificity scores were generally high and closely aligned, with the KNN algorithm emerging as the top performer in predicting non-diabetes cases among middle-aged adults, boasting a specificity score of 97.83%. Notably, the RF model excelled in overall accuracy and AUC, achieving scores of 96.58% and 96.00%, respectively. This aligns with the consensus in previous studies, highlighting the effectiveness of ensemble ML techniques in predicting or diagnosing diabetes across diverse populations.[42,44,47-49]

This study has limitations, including its reliance on the dataset from India, which may affect the generalizability of results due to regional differences in healthcare and lifestyle. Additionally, given the secondary nature of our study's data, there was no information on the type of diabetes the patient had. As such, our study assumes that all diabetes types share the examined symptoms, which may not be universally applicable. This potentially limits the clinical implications of the prevalent diabetes type in the dataset.

Conclusion

In conclusion, our study reveals key symptoms associated with diabetes in middle-aged adults, such as polyuria, polydipsia, gender, alopecia, irritability, and sudden weight loss. Additionally, it highlights the effectiveness of ML models, like the RF classifier, in predicting diabetes among middle-aged adults. Future studies should investigate genetic predispositions, lifestyle factors, and environmental influences to refine predictive models and develop personalized healthcare strategies, aiming to improve outcomes for at-risk individuals in this age group.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Roglić G. WHO global report on diabetes: A summary. Int J Noncommun Dis 2016;1:3.
2. Chou CY, Hsu D, Chou CH. Predicting the onset of diabetes with machine learning methods. J Pers Med 2023;13:406.
3. Centers for Disease Control and Prevention. Diabetes Basics. 2024. Available from: https://www.cdc.gov/diabetes/about/index.html. [Last accessed on 7 Sep 2024].
4. World Health Organization (WHO). Diabetes. 2021. Available from: https://www.who.int/news-room/facts-in-pictures/detail/diabetes. [Last accessed on 16 Oct 2023].
5. Maahs DM, West NA, Lawrence JM, Mayer-Davis EJ. Epidemiology of type 1 diabetes. Endocrinol Metab Clin North Am 2010;39:481-97.
6. Dariya B, Chalikonda G, Srivani G, Alam A, Nagaraju GP. Pathophysiology, etiology, epidemiology of type 1 diabetes and computational approaches for immune targets and therapy. Crit Rev Immunol 2019;39:239-65.
7. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2010;33:S62-9.
8. Mobasseri M, Shirmohammadi M, Amiri T, Vahed N, Fard HH, Ghojazadeh M. Prevalence and incidence of type 1 diabetes in the world: A systematic review and meta-analysis. Health Promot Perspect 2020;10:98-115.
9. Gale EAM. The rise of childhood type 1 diabetes in the 20th century. Diabetes 2002;51:3353-61.
10. Meigs JB, Muller DC, Nathan DM, Blake DR, Andres R; Baltimore Longitudinal Study of Aging. The natural history of progression from normal glucose tolerance to type 2 diabetes in the Baltimore Longitudinal Study of Aging. Diabetes 2003;52:1475-84.
11. Zeng B, Lu Y, Hajifathalian K, Bentham J, Di Cesare M, Danaei G, *et al*. Worldwide trends in diabetes since 1980: A pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016;387:1513-30.
12. Ramachandran A. Know the signs and symptoms of diabetes. Indian J Med Res 2014;140:579-81.

13. Plows JF, Stanley JC, Baker PN, Reynolds CM, Vickers MH. The pathophysiology of gestational diabetes mellitus. Int J Mol Sci 2018;19:3342.
14. Juan J, Yang H. Prevalence, prevention, and lifestyle intervention of gestational diabetes mellitus in China. Int J Environ Res Public Health 2020;17:9517.
15. Blair ME. Diabetes mellitus review. Urol Nurs 2016;36:27.
16. World Health Organization (WHO). Diabetes. 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/diabetes. [Last accessed on 17 Oct 2023].
17. Emancipator K. Laboratory diagnosis and monitoring of diabetes mellitus. Am J Clin Pathol 1999;112:665-74.
18. Sacks DB, Bruns DE, Goldstein DE, Maclaren NK, McDonald JM, Parrott M. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Clin Chem 2002;48:436-72.
19. Kaur G, Lakshmi PVM, Rastogi A, Bhansali A, Jain S, Teerawattananon Y, *et al*. Diagnostic accuracy of tests for type 2 diabetes and prediabetes: A systematic review and meta-analysis. PLoS One 2020;15:e0242415.
20. Man B, Schwartz A, Xia Y, Gerber BS. Individualized diabetes risk prediction in women with a history of gestational diabetes. Diabetes 2018;67:1293.
21. Li T, Quan H, Zhang H, Lin L, Lü L, Ou Q, *et al*. Type 2 diabetes is more predictable in women than men by multiple anthropometric and biochemical measures. Sci Rep 2021;11:6062.
22. Zhang X, Zhao X, Huo L, Yuan N, Sun J, Du J, *et al*. Risk prediction model of gestational diabetes mellitus based on nomogram in a Chinese population cohort study. Sci Rep 2020;10:21223.
23. Boutilier JJ, Chan TCY, Ranjan M, Deo S. Risk stratification for early detection of diabetes and hypertension in resource-limited settings: Machine learning analysis. J Med Internet Res 2021;23:e20123.
24. Fernández-Edreira D, Liñares-Blanco J, Fernández-Lozano C. Machine learning analysis of the human infant gut microbiome identifies influential species in type 1 diabetes. Expert Syst Appl 2021;185:115648.
25. Kushwaha S, Srivastava RN, Jain R, Sagar V, Aggarwal AK, Bhadada SK, *et al*. Harnessing machine learning models for non-invasive pre-diabetes screening in children and adolescents. Comput Methods Programs Biomed 2022;226:107180.
26. Hu H, Lai T, Farid F. Feasibility study of constructing a screening tool for adolescent diabetes detection applying machine learning methods. Sensors 2022;22:6155.
27. Diabetes prediction in teenagers using machine learning algorithms. IEEE Conference Publication, IEEE Xplore; 2023. Available from: https://ieeexplore.ieee.org/abstract/document/10112286. [Last accessed on 17 Oct 2023].
28. Islam MM, Ferdousi R, Rahman S, Bushra HY. Likelihood prediction of diabetes at early stage using data mining techniques. Adv Intell Syst Comput 2019;992:113-25.
29. Joseph LP. Diabetes datasets. Mendeley Data. 2022. Available from: https://doi.org/10.17632/7zcc8v6hvp.1 [Last accessed on 15 Sep 2023].
30. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw 2010;36:1-13.
31. Devi KBL, Meitei KT, Singh SD. Prevalence of type 2 diabetes and its signs and symptoms among the Meiteis of Manipur, India. J Anthropol Surv India 2022;72:59-70.
32. Kumar A, Kaplowitz PB. Patient age, race and the type of diabetes have an impact on the presenting symptoms, latency before diagnosis and laboratory abnormalities at time of diagnosis of diabetes mellitus in children. J Clin Res Pediatr Endocrinol 2009;1:227-32.
33. Pawar SD, Thakur P, Radhe B, Jadhav H, Behere V, Pagar V. The accuracy of polyuria, polydipsia, polyphagia, and Indian Diabetes Risk Score in adults screened for diabetes mellitus type-II. Med J Dr DY Patil University 2017;10:263.
34. Pawar SD, Naik JD, Prabhu PM, Jatti GM, Jadhav S, Radhe B. Comparative evaluation of Indian Diabetes Risk Score and Finnish Diabetes Risk Score for predicting risk of diabetes mellitus type II: A teaching hospital-based survey in Maharashtra. J Family Med Prim Care 2017;6:120-5.
35. Boles A, Kandimalla R, Reddy PH. Dynamics of diabetes and obesity: Epidemiological perspective. Biochim Biophys Acta Mol Basis Dis 2017;1863:1026-36.
36. Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, *et al*. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA 2003;289:76-9.
37. Barnes AS. The epidemic of obesity and diabetes: Trends and treatments. Tex Heart Inst J 2011;38:142-4.
38. Fahey TJ, Sadaty A, Jones WG, Barber A, Smoller BR, Shires GT. Diabetes impairs the late inflammatory response to wound healing. J Surg Res 1991;50:308-13.
39. Tang Y, Zhang MJ, Hellmann J, Kosuri M, Bhatnagar A, Spite M. Proresolution therapy for the treatment of delayed healing of diabetic wounds. Diabetes 2013;62:618-27.
40. Kumari S, Kumar D, Mittal M. An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. Int J Cognit Comput Eng 2021;2:40-6.
41. Viswanatha V, Ramachandra AC, Murthy D, Thanishka. Diabetes prediction using machine learning approach. Strad Res 2023;10:75-82.
42. Sadhu A, Jadli A. Early-stage diabetes risk prediction: A comparative analysis of classification algorithms. Int Adv Res J Sci Eng Technol 2021;8:193-201.
43. Xue J, Min F, Ma F. Research on diabetes prediction method based on machine learning. J Phys 2020;1684:012062.
44. Le TM, Vo TM, Pham TT, Dao SVT. A novel wrapper–based feature selection for early diabetes prediction enhanced with a metaheuristic. IEEE Access 2021;9:7869-84.
45. Julius AO, Ayokunle AO, Ibrahim FO. Early diabetic risk prediction using machine learning classification techniques. Int J Innov Sci Res Technol 2021;6:502.
46. Hassan AS, Malaserene I, Leema AA. Diabetes mellitus prediction using classification techniques. Int J Innov Technol Explor Eng 2020;9:2080-4.
47. Birjais R, Mourya AK, Chauhan R, Kaur H. Prediction and diagnosis of future diabetes risk: A machine learning approach. SN Appl Sci 2019;1:1-8.
48. Phongying M, Hiriote S. Diabetes classification using machine learning techniques. Computation 2023;11:96.
49. Maniruzzaman M, Rahman MJ, Ahammed B, Abedin MM. Classification and prediction of diabetes disease using machine learning paradigm. Health Inform Sci Syst 2020;8:1-14.