engineering proceedings

Proceeding Paper

Application of Machine Learning for Predicting Diabetes Mellitus Using Body Composition Parameters †

Elena Vizitiu Lakhdari [1], Claudiu Cobuz [2], Adriana Slusar [3] and Dan Milici [4,*]

1 Interdisciplinary Research Center for Motor Science and Human Health, "Ștefan cel Mare" University of Suceava, Universitatii Street, 13, 720229 Suceava, Romania; elenav@usm.ro
2 Faculty of Medicine and Biological Sciences, "Ștefan cel Mare" University of Suceava, Universitatii Street, 13, 720229 Suceava, Romania; claudiu.cobuz@usm.ro
3 Faculty of Physical Education and Sport, "Ștefan cel Mare" University of Suceava, Universitatii Street, 13, 720229 Suceava, Romania; slusaradriana87@gmail.com
4 Faculty of Electrical Engineering and Computer Science, "Ștefan cel Mare" University of Suceava, Universitatii Street, 13, 720229 Suceava, Romania
* Correspondence: dam@usm.ro; Tel.: +40-741221406
† Presented at the International Conference on Electromagnetic Fields, Signals and BioMedical Engineering (ICEMS-BIOMED), Suceava, Romania, 7–9 May 2026.

Abstract

Diabetes mellitus represents a major public health issue, and the early identification of individuals at risk is essential for prevention. The aim of this study was to develop a predictive model for identifying diabetes mellitus based on body composition parameters and anthropometric characteristics, using a logistic regression model as a baseline machine learning approach. The study, designed as a cross-sectional observational investigation, included 81 adult participants assessed through anthropometric measurements and bioimpedance analysis. The model achieved an accuracy of 84%, sensitivity of 90%, specificity of 74%, precision of 85%, and an AUC of 0.897. The results highlight the contribution of diet, physical activity, age, and BMI in estimating the risk of diabetes mellitus.

Keywords: artificial intelligence; diabetes prediction; body composition; machine learning; metabolic risk; anthropometric parameters; health assessment; predictive modeling

1. Introduction

Diabetes mellitus is one of the most significant global public health issues, being associated with major metabolic and cardiovascular complications. The prevalence of this condition is continuously increasing, with estimates indicating that it may reach approximately 642 million cases by 2040 [1]. This chronic metabolic disease, characterized by hyperglycemia resulting from defects in insulin secretion or action, is associated with a significant increase in morbidity and mortality, particularly due to cardiovascular complications [2]. Additionally, diabetes increases the risk of developing hypertension and other chronic conditions [3]. An important complication is diabetic autonomic neuropathy, which can significantly reduce quality of life and requires early identification for effective intervention [4]. Early identification of individuals at risk is essential for preventing disease progression. However, body mass index (BMI) does not always reflect the true metabolic risk, as some individuals with elevated BMI do not exhibit increased metabolic risk. At the same time, factors such as visceral adiposity and sarcopenia may increase risk even in individuals with normal body weight [5]. In this context, body composition assessment becomes an essential tool in the analysis of metabolic risk. Nevertheless, traditional methods

Academic Editors: Mihai Munteanu, Simona Miclăuș, Paul Bechet and Radu Ciorap

Published: 30 June 2026

 https://doi.org/10.3390/engproc2026148003

present methodological and clinical applicability limitations [6]. For this reason, the use of modern non-invasive technologies is gaining increasing importance [7]. At the same time, interest in the use of artificial intelligence in medicine has grown significantly. AI enables the analysis of large volumes of data and the development of efficient predictive models. It can be defined as the ability of computer systems to perform tasks that normally require human intelligence [8]. The application of artificial intelligence methods in diabetology facilitates early risk identification and the personalization of therapeutic interventions [9]. Furthermore, the concept of predictive health promotes the transition from reactive to proactive medicine through the integration of physiological data into predictive models [10]. Recent studies have demonstrated that the regional distribution of adipose tissue plays an important role in the development of diabetes, with age, fat mass, and its distribution in the limbs and trunk identified as relevant predictors [5]. Artificial intelligence has also been successfully used in the prediction of type 1 diabetes, highlighting the role of neonatal and perinatal factors [11]. Diabetes remains a condition with a major global impact, requiring effective prevention and management strategies based on advanced technologies [12,13]. In this context, the aim of the present study was to develop a predictive model for identifying patients with diabetes mellitus based on body composition parameters and anthropometric characteristics, using artificial intelligence methods.

2. Materials and Methods

2.1. Study Design and Participants

The conducted study is an observational, cross-sectional investigation based on the analysis of anthropometric parameters and body composition. The study sample included 81 adult participants, aged between 21 and 87 years. Data were collected at the Outpatient Clinic for Diabetes Mellitus, Nutrition, and Metabolic Diseases within the Suceava County Clinical Hospital, through anthropometric measurements and body composition assessment using the bioimpedance analyzer in Body Europe B.V., Hessenbergweg 75 A, 1101CX, Amsterdam, The Netherlands.

2.2. Variables

The analyzed variables included the following: demographic data (age and sex); anthropometric parameters (height, weight, and body mass index-BMI); body composition parameters (muscle mass, adipose tissue, body water, and bone tissue); lifestyle variables (level of physical activity and diet, assessed using a categorical scale based on self-reported dietary patterns); the dependent variable: diabetes mellitus (0 = absent, 1 = present).

2.3. Data Processing

The database was organized in a tabular format. The following operations were performed: standardization of categorical variables (0/1); transformation of numerical values into decimal format; identification of extreme values (outliers); imputation of missing values based on BMI, sex, and the distribution of the existing data.

2.4. Statistical and AI Model

A logistic regression model was used for the prediction of diabetes mellitus, serving as a baseline machine learning approach, as it is appropriate for binary dependent variables.

The model function is:

$$P(Y = 1) = \frac{1}{1 + e^{-z}} \tag{1}$$

where z represents the linear combination of the independent variables.

The linear model is defined as follows:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \quad (2)$$

The model coefficients were estimated using the maximum likelihood method. The Solver tool in Microsoft Excel was used to minimize the Log Loss function; however, this approach may present limitations in terms of reproducibility compared with standard statistical software:

$$\text{LogLoss} = -[y\ln(p) + (1-y)\ln(1-p)] \quad (3)$$

2.5. Model Evaluation

The model performance was evaluated using the following indicators: accuracy, sensitivity, specificity, precision, F1-score, and AUC. To enhance model robustness, internal validation considerations were taken into account during the analysis. However, due to the exploratory nature of the study and the limited sample size, external validation and cross-validation procedures were not implemented.

3. Results

The logistic regression model demonstrated satisfactory performance, achieving an accuracy of 84%. The model's sensitivity was 90%, and its specificity was 74%. Precision was 85%, and the F1-score reached 0.87.

The AUC value was 0.897, indicating a very good discriminative ability of the model.

3.1. Confusion Matrix

The confusion matrix indicates good model performance. A total of 45 positive cases and 23 negative cases were correctly identified, highlighting a good discriminative performance. The low number of false positives (5) and false negatives (8) suggests an appropriate balance between sensitivity and specificity (see Table 1).

Table 1. Confusion matrix for the logistic regression model.

Type	Value
TP	45
TN	23
FP	8
FN	5

3.2. Logistic Regression Coefficients

The coefficients of the logistic regression model provide information regarding the direction and strength of the relationship between the independent variables and the probability of developing diabetes mellitus. The negative intercept (−3.51) indicates a low baseline probability of diabetes occurrence in the absence of the analyzed factors. The variable sex shows a relatively high positive coefficient (1.72), suggesting that the category coded as 1 is associated with a significantly higher probability of developing diabetes. Age has a positive coefficient of 0.05, indicating a gradual increase in risk with advancing age. BMI shows a positive coefficient of 0.12, confirming that higher BMI values are associated with an increased risk of diabetes mellitus. Muscle mass has a small positive coefficient of 0.07, suggesting a weak influence on the probability of developing the disease. Adipose tissue presents a negative coefficient of −0.13, a result that should be interpreted with caution, as it may be influenced by the interdependence of body composition variables included in the model. In contrast, body water has a more pronounced negative coefficient of

−0.61, indicating a significant protective effect associated with a more favorable metabolic profile. Bone tissue shows a positive coefficient of 0.45, suggesting a moderate association with the probability of developing diabetes, possibly reflecting indirect relationships with other anthropometric parameters. The level of physical activity has a negative coefficient of −0.91, highlighting a significant protective effect, consistent with the role of physical activity in diabetes prevention. In contrast, diet shows the highest positive coefficient (2.30), indicating a major influence on diabetes risk, depending on the type of diet analyzed. The results emphasize that lifestyle factors, particularly diet and physical activity, along with anthropometric parameters such as BMI and age, contribute to the estimation of diabetes mellitus risk (see Table 2).

Table 2. Logistic regression coefficients.

Variable	Coefficient
Intercept	−3.51
Sex	1.72
Age	0.05
BMI	0.12
Muscle mass	0.07
Adipose tissue	−0.13
Body water	−0.61
Bone tissue	0.45
Physical activity	−0.91
Diet	2.30

3.3. Odds Ratio Interpretation

Odds ratios were calculated to assess the magnitude of the association between each predictor and the probability of developing diabetes. The results showed that diet had the strongest influence (OR = 9.97), followed by sex (OR = 5.59), indicating a substantial increase in the likelihood of diabetes for these variables. Physical activity demonstrated a significant protective effect (OR = 0.40), while body water percentage was also inversely associated with diabetes risk (OR = 0.54). Age and body mass index showed moderate positive associations (OR = 1.05 and OR = 1.13, respectively), suggesting a gradual increase in risk. Although adipose tissue showed a negative association (OR = 0.88), this result should be interpreted with caution due to potential multicollinearity among body composition variables.

4. Discussion

Artificial intelligence-based analysis of abdominal CT images enables the assessment of body composition and the prediction of diabetes risk [14]. Studies have shown that the regional distribution of adipose tissue influences diabetes risk, with the most important predictors being age, fat mass and fat percentages in the limbs and trunk [15]. Additionally, fat-free mass has been negatively associated with diabetes, while XGBoost models have demonstrated high predictive performance [16]. The results of the present study are consistent with these trends, with the logistic regression model showing a good capacity for diabetes prediction. Lifestyle factors, particularly diet and physical activity, had the strongest influence on the probability of disease occurrence, while anthropometric parameters such as BMI and age were positively associated with diabetes risk. In contrast, body water and physical activity showed a protective effect. The interpretation of the coefficients should be made in the context of the interdependence of body composition variables, which may influence the direction of the observed associations.

The limitations of the study include the relatively small sample size, the cross-sectional design, which does not allow causal inference, and the absence of cross-validation and external validation, which may affect model generalizability. Additionally, the relatively small sample size limits the strength of predictive conclusions. Future studies should include larger samples and validate the models on independent datasets to confirm the accuracy of the obtained results. Despite these limitations, the model provides a transparent and clinically applicable baseline framework for diabetes risk prediction.

Author Contributions: Conceptualization, E.V.L. and D.M.; methodology, E.V.L.; validation, C.C., A.S. and E.V.L.; formal analysis, E.V.L.; investigation, C.C.; resources, A.S.; writing—original draft preparation, C.C.; writing—review and editing, A.S.; supervision, E.V.L.; project administration, D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ML	Machine Learning
DM	Diabetes Mellitus
BMI	Body Mass Index
OR	Odds Ratio
CI	Confidence Interval
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
F1-score	F1 Score
LogLoss	Logarithmic Loss

References

1. Nematollahi, M.A.; Joloudari, J.H.; Zare, O.; Maftoun, M.; Shadkam, N.; Sharifrazi, D.; Alizadehsani, R.; Asadollahi, A. Evolution of diabetes prediction using fusion of ANOVA, ADASYN technique and XGBoost. *J. Diabetes Metab. Disord.* **2025**, *24*, 151. [CrossRef] [PubMed]
2. Nath, T.; Ahima, R.S.; Santhanam, P. Body fat predicts exercise capacity in persons with type 2 diabetes mellitus: A machine learning approach. *PLoS ONE* **2021**, *16*, e0248039. [CrossRef] [PubMed]
3. Khan, S.; Shah, Z. Artificial intelligence–based diabetes risk prediction from longitudinal DXA bone measurements. *Sci. Rep.* **2025**, *15*, 25706. [CrossRef] [PubMed]
4. Toderean, R.; Cobuz, M.; Dimian, M.; Cobuz, C. From evaluation to prediction: Analysis of diabetic autonomic neuropathy using Sudoscan and artificial intelligence. *Appl. Sci.* **2024**, *14*, 7406. [CrossRef]
5. Rigla, M.; García-Sáez, G.; Pons, B.; Hernando, M.E. Artificial intelligence methodologies and their application to diabetes. *J. Diabetes Sci. Technol.* **2018**, *12*, 303–310. [CrossRef] [PubMed]
6. Cobuz, C.; Silișteanu, S.C.; Cobuz, M.; Costea, A.I.; Lakhdari, E.V. Preliminary analysis of patient assessment within a study on the impact of hydrokinetotherapy on body composition and metabolic disease risk. *Balneo PRM Res. J.* **2024**, *15*, 752. [CrossRef]

7. Cobuz, C.; Ungureanu-Iuga, M.; Anton-Paduraru, D.T.; Cobuz, M. Possible use of the Sudoscan nephropathy risk score in chronic kidney disease diagnosis. *Biosensors* **2025**, *15*, 620. [CrossRef] [PubMed]
8. Santhanam, P.; Nath, T.; Peng, C.; Bai, H.; Zhang, H.; Ahima, R.S.; Chellappa, R. Artificial intelligence and body composition. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2023**, *17*, 102732. [CrossRef] [PubMed]
9. Yesupatham, A.; Das, R.; Bharani, G.; Shaikmeeran, M.; Saraswathy, R. Artificial intelligence model as a tool to predict prediabetes. *Sci. Rep.* **2025**, *15*, 43421. [CrossRef] [PubMed]
10. Boschiero, D.; Gallotta, A.; Ferrari, F.; Dragoumani, K.; Lamprou, G.; Vlachakis, D.; Chrousos, G.P. Predictive health index: Integrating body composition and heart rate variability metrics with artificial intelligence to predict chronic disease risk. *Hormones* **2026**, *25*, 181–193. [CrossRef] [PubMed]
11. Cobuz, C.; Ungureanu-Iuga, M.; Cobuz, M. Neonatal and birth risk factors for type 1 diabetes mellitus: Prediction using an artificial neural network. *Life* **2025**, *15*, 1800. [CrossRef] [PubMed]
12. Alam, M.A.; Sohel, A.; Hasan, K.M.; Islam, M.A. Machine learning and artificial intelligence in diabetes prediction and management: A comprehensive review of models. *J. Next-Gen Eng. Syst.* **2024**, *1*, 107–124. [CrossRef]
13. Anghel, L.; Cobuz, C.; Benchea, L.-C.; Maciuc, V.; Cobuz, M.; Sascău, R.-A.; Stătescu, C. Sudomotor dysfunction as an early marker of autonomic and cardiovascular risk in diabetes. *Biosensors* **2025**, *15*, 372. [CrossRef] [PubMed]
14. Kim, Y.H.; Yoon, J.W.; Lee, B.H.; Yoon, J.H.; Choe, H.J.; Oh, T.J.; Lee, J.M.; Cho, Y.M. Artificial intelligence-based body composition analysis using CT images predicts diabetes. *J. Diabetes Investig.* **2025**, *16*, 272–284. [CrossRef] [PubMed]
15. Nematollahi, M.A.; Askarinejad, A.; Asadollahi, A.; Salimi, M.; Moghadami, M.; Sasannia, S.; Sheikh, M.S.I. Association and predictive capability of body composition and diabetes mellitus using artificial intelligence. *Res. Sq.* **2022**. [CrossRef] [PubMed]
16. Nematollahi, M.A.; Askarinejad, A.; Asadollahi, A.; Bazrafshan, M.; Sarejloo, S.; Moghadami, M.; Sasannia, S.; Farjam, M.; Homayounfar, R.; Pezeshki, B.; et al. A cohort study on the predictive capability of body composition for diabetes mellitus using machine learning. *J. Diabetes Metab. Disord.* **2024**, *23*, 773–781. [CrossRef] [PubMed]