TOSHKENT TIBBIYOT AKADEMIYASI URGANCH FILIALI
JANUBIY OROLBO'YI TIBBIYOT JURNALI
1-TOM, 4-SON. 2025
14.00.00 - TIBBIYOT FANLARI ISSN: 3093-8740

UDK: 615.3-09

COMPARATIVE ANALYSIS OF PREDICTION MODELS FOR DIABETES BASED ON ANTHROPOMETRIC INDICATORS

Mukhamedova Vazira Mukimovna
Docent, Department of Hospital Therapy and Endocrinology
Andijan State Medical Institute
dr.vazira12345@gmail.com
ORCID : 0000-0003-1022-2217

Abstract

The prevalence of type 2 diabetes mellitus (T2DM) continues to increase worldwide, emphasizing the need for early detection tools that are accurate, accessible, and cost-effective. Anthropometric indicators—such as body mass index (BMI), waist circumference (WC), waist-to-hip ratio (WHR), waist-to-height ratio (WHtR), and body fat percentage—are widely used predictors of metabolic risk. Numerous statistical and machine learning models have been developed to forecast diabetes risk using these indicators. This study provides a comparative analysis of traditional regression-based models and modern machine learning algorithms to determine their predictive performance using anthropometric data.

Keywords: diabetes prediction, anthropometry, BMI, waist circumference, machine learning, risk assessment.

Annotatsiya

2-tip qandli diabet (T2DM) dunyo bo‘ylab ortib borayotgan kasallik bo‘lib, uni erta aniqlash uchun aniq, qulay va iqtisodiy jihatdan samarali usullarga ehtiyoj ortib bormoqda. Tana massasi indeksi (BMI), bel aylanasining uzunligi (WC), bel-son nisbati (WHR), bel-bo‘y nisbati (WHtR) hamda tana yog‘i foizi kabi antropometrik ko‘rsatkichlar metabolik xavfni baholashda keng qo‘llaniladi. Ushbu ko‘rsatkichlar asosida qandli diabet xavfini bashorat qilish uchun ko‘plab statistik modellar va mashina o‘rganish algoritmlari ishlab chiqilgan. Mazkur tadqiqot an’anaviy regressiyaga asoslangan modellar va zamonaviy mashina o‘rganish algoritmlarining antropometrik ma’lumotlar asosidagi bashorat samaradorligini solishtirma tahlil qiladi.

Kalit so‘zlar: diabetni bashorat qilish, antropometriya, BMI, bel aylanasining uzunligi, mashina o‘rganish, xavfni baholash.

Аннотация

Распространённость сахарного диабета 2-го типа (СД2) продолжает увеличиваться во всём мире, что подчёркивает необходимость ранних методов выявления, отличающихся точностью, доступностью и экономической эффективностью. Антропометрические показатели — индекс массы тела (ИМТ), окружность талии (ОТ), отношение талии к бёдрам (ОТ/ОБ), отношение талии к росту (ОТ/Р), а также процент жировой массы тела — широко используются как предикторы метаболического риска. На основе этих показателей разработано множество статистических моделей и алгоритмов машинного обучения для прогнозирования риска диабета. В данном исследовании проводится сравнительный анализ традиционных регрессионных моделей и современных алгоритмов машинного обучения с целью определения их прогностической эффективности при использовании антропометрических данных.

Ключевые слова: прогнозирование диабета, антропометрия, ИМТ, окружность талии, машинное обучение, оценка риска.

Introduction

The global rise in type 2 diabetes mellitus (T2DM) continues to pose a major public health challenge, with prevalence increasing rapidly in both developed and developing countries [1]. Early identification of individuals at high risk is critical for reducing long-term complications and healthcare burden. Anthropometric indicators—such as body mass index (BMI), waist circumference (WC), waist-to-hip ratio (WHR), and waist-to-height ratio (WHtR)—are widely used as cost-effective, non-invasive measures for assessing obesity-related metabolic risk [2,3]. Numerous epidemiological studies indicate that central obesity, rather than general adiposity, is a stronger predictor of insulin resistance and subsequent diabetes development [4,5]. As a result, anthropometric parameters have become integral components of diabetes screening frameworks in many clinical guidelines [6].

Traditional statistical risk assessment models such as the Framingham Diabetes Risk Score and the Finnish Diabetes Risk Score rely heavily on anthropometric and lifestyle variables to estimate diabetes probability [7]. However, these models are constrained by linear assumptions and may fail to capture complex interactions among anthropometric measures, metabolic biomarkers, and demographic characteristics [8]. With the emergence of large-scale health datasets and computational advancements, machine learning approaches have gained prominence for diabetes prediction because of their ability to model nonlinear relationships with higher accuracy [9,10].

Recent studies comparing anthropometry-based prediction algorithms demonstrate that WC and WHtR outperform BMI as predictors of metabolic dysfunction due to their ability to reflect visceral fat accumulation, a key determinant of insulin sensitivity [11,12]. Evidence suggests that WHtR may serve as a universally applicable predictor across ethnic groups, genders, and age categories, outperforming BMI in several population-based cohorts [13]. Moreover, combining multiple anthropometric indicators within predictive models improves sensitivity and specificity, especially when assessing heterogeneous populations [14,15].

Machine learning techniques—including random forests, support vector machines, gradient boosting models, and artificial neural networks—have been applied to anthropometric datasets with promising results [16]. These algorithms can automatically detect subtle, high-dimensional patterns that traditional regression models typically overlook. Studies using integrated anthropometric-clinical datasets show significant improvements in area-under-curve performance metrics, demonstrating the added predictive value of machine learning approaches [17,18].

Additionally, the expanding availability of mobile health technologies, wearable devices, and community-based health surveys has facilitated continuous collection of anthropometric and lifestyle data, creating opportunities for real-time diabetes risk prediction models [19]. These advances are aligned with global strategies emphasizing early detection and preventive interventions to reduce T2DM incidence [20].

Given the increasing role of anthropometric indicators in diabetes risk assessment and the growing diversity of computational prediction models, a systematic comparison of traditional and machine learning-based approaches is essential. Such an analysis can guide clinicians, public health practitioners, and policymakers in selecting appropriate, population-specific tools for early diabetes detection. The present study aims to provide a comprehensive comparative evaluation of these models, highlighting the strengths, limitations, and practical implications of anthropometry-based diabetes prediction systems.

Materials and Methods

TOSHKENT TIBBIYOT AKADEMIYASI URGANCH FILIALI
JANUBIY OROLBO'YI TIBBIYOT JURNALI
1-TOM, 4-SON. 2025
14.00.00 - TIBBIYOT FANLARI ISSN: 3093-8740

This study employed a comprehensive methodological approach to compare prediction models that utilize anthropometric indicators for assessing the risk of type 2 diabetes mellitus. The analysis was based on an extensive review of scientific literature aimed at identifying the effectiveness, strengths, and limitations of both traditional statistical models and modern machine learning algorithms applied to anthropometric data. Relevant studies were identified through systematic searches conducted across major scientific databases, including PubMed, Scopus, Web of Science, IEEE Xplore, and Google Scholar. The literature search covered publications from 2000 to 2024 and included population-based cohort studies, clinical datasets, and cross-sectional analyses that reported anthropometric measurements alongside diabetes-related outcomes. Studies were included if they incorporated at least one anthropometric predictor, reported incidence or diagnostic confirmation of diabetes, employed a predictive modeling approach, and provided sufficient performance metrics for comparative evaluation. Studies lacking clear methodological descriptions or relying exclusively on biochemical markers were excluded.

Anthropometric variables extracted from the reviewed studies included body mass index, waist circumference, waist-to-hip ratio, waist-to-height ratio, body fat percentage, and visceral fat estimates obtained through bioimpedance or imaging techniques. These indicators were selected because they represent biologically meaningful measures of adiposity and fat distribution, which are strongly associated with metabolic risk. Their interpretation followed standardized measurement guidelines established by the World Health Organization to ensure consistency across studies.

Prediction models discussed in the literature were broadly categorized into traditional statistical approaches and machine learning-based algorithms. Statistical models such as logistic regression, generalized linear models, and discriminant analysis were frequently used to explore linear relationships between anthropometric predictors and diabetes risk. These models offered interpretability and ease of application but were limited in their ability to capture nonlinear interactions or high-dimensional patterns within the data. In contrast, machine learning algorithms—including decision trees, random forests, support vector machines, k-nearest neighbors, gradient boosting techniques, and artificial neural networks—were designed to detect complex, nonlinear associations and interactions among predictors, potentially improving predictive accuracy.

Most studies trained their models using a training subset comprising 70–80 percent of available data, reserving the remaining portion for validation. Cross-validation procedures, typically involving 5-fold or 10-fold techniques, were widely employed to assess model robustness and reduce overfitting. Machine learning models often required hyperparameter optimization, which was commonly performed using grid search or randomized search procedures. Performance assessment relied on metrics such as accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve, with the latter serving as the primary indicator of discriminative capacity. Calibration measures were additionally used in some studies to evaluate the agreement between predicted and observed risks.

Comparative evaluation considered not only predictive performance but also interpretability, generalizability, and clinical applicability. Statistical models were recognized for their simplicity and transparency, whereas machine learning models demonstrated superior predictive performance but sometimes lacked interpretability. Particular attention was given to studies that integrated multiple anthropometric indicators, as multivariable approaches often produced more accurate predictions than models relying on single indicators. Differences in model performance across demographic subgroups—including variations by age, sex, and ethnicity—were also analyzed to assess the adaptability of anthropometric prediction models across diverse populations.

Because this study was based exclusively on published literature and secondary data, no direct involvement of human subjects occurred, and ethical approval was not required. The studies included in this review were assumed to have been conducted in compliance with ethical research standards as reported by their authors. Through this methodological framework, the comparative analysis achieved a thorough and academically rigorous evaluation of existing diabetes prediction models based on anthropometric indicators.

Results

The comparative analysis of prediction models based on anthropometric indicators demonstrated consistent patterns across the reviewed literature regarding the relative predictive strength of different anthropometric measures and the performance of various modeling approaches. Among single anthropometric predictors, waist circumference and waist-to-height ratio emerged as the most robust indicators of future diabetes risk. These measures showed stronger correlations with incident diabetes than body mass index, largely due to their ability to capture central adiposity and visceral fat accumulation, which are more closely associated with insulin resistance. Studies consistently reported that models relying solely on body mass index yielded moderate predictive accuracy and were insufficient for accurately identifying individuals with central obesity who remain at elevated metabolic risk despite having a normal or mildly elevated body mass index.

Multivariate models combining several anthropometric indicators demonstrated improved discriminatory performance compared with single-indicator models. When waist circumference, waist-to-height ratio, or waist-to-hip ratio were included in regression frameworks, predictive accuracy increased substantially, often achieving area-under-curve values above 0.75. These findings suggest that integrating complementary anthropometric dimensions—reflecting both total and regional adiposity—enhances the ability of prediction models to detect early metabolic disturbances.

Machine learning models outperformed traditional statistical approaches in nearly all comparative studies. Algorithms such as random forests, gradient boosting machines, and support vector machines consistently achieved higher accuracy, sensitivity, and area-under-curve values. Many machine learning models reported area-under-curve metrics ranging from 0.80 to 0.90, demonstrating their ability to capture nonlinear effects and subtle interactions between anthropometric variables. Random forest models frequently ranked among the highest-performing approaches due to their inherent capability to handle variable interactions and reduce overfitting through ensemble learning. Gradient boosting models exhibited similar or superior performance when trained on sufficiently large datasets.

Neural network models also showed strong predictive capabilities, particularly in studies incorporating both anthropometric indicators and demographic variables such as age, sex, and family history. Their performance improved markedly when datasets contained high sample sizes and when input variables reflected diverse dimensions of metabolic health. However, the complexity and relatively low interpretability of neural networks were noted as potential limitations for clinical adoption.

The comparative findings further revealed that model performance varied across populations, with certain anthropometric predictors demonstrating higher predictive utility in specific ethnic groups. Waist-to-height ratio performed well across nearly all populations, whereas waist-to-hip ratio exhibited variable performance depending on gender and ethnic background. These differences indicate that population-specific calibration may be necessary to optimize prediction accuracy.

Overall, the analysis demonstrated that anthropometric indicators—particularly waist-related measures—are strong predictors of diabetes risk, and that their predictive value is substantially enhanced when incorporated into machine learning algorithms that can model nonlinear patterns. The superiority of machine learning approaches suggests that they hold considerable promise for the development of accurate and scalable diabetes prediction tools suitable for population-level screening.

Discussion

The findings of this comparative analysis highlight the importance of central obesity indicators—especially waist circumference and waist-to-height ratio—as key predictors of diabetes risk. The strong performance of these measures reflects their ability to capture visceral adiposity, which is more closely linked to insulin resistance than general body weight.

Traditional regression models remain useful due to their simplicity, interpretability, and ease of clinical application. However, their predictive capacity is limited by linear assumptions and reduced ability to model complex risk patterns.

Machine learning algorithms overcome these limitations by integrating nonlinear relationships, interaction effects, and high-dimensional data structures. Their higher predictive performance suggests that incorporating AI-based tools into screening programs may facilitate earlier detection of diabetes and more personalized preventive interventions.

Despite these advantages, machine learning systems also face challenges. They often require large datasets for optimal performance, may lack interpretability, and depend on computational resources not always accessible in low-resource settings. Therefore, hybrid approaches combining interpretability of regression models with the predictive strength of machine learning may represent an optimal solution.

Conclusion

The findings of this comparative analysis demonstrate that anthropometric indicators represent essential and highly accessible tools for predicting the risk of type 2 diabetes mellitus. Among the various measures examined, waist circumference and waist-to-height ratio consistently emerged as the strongest standalone predictors, underscoring the central role of visceral adiposity in the development of metabolic dysfunction. While body mass index remains widely used in clinical practice, its limitations in distinguishing between fat and lean mass or detecting central obesity reduce its effectiveness as a reliable predictor when used in isolation. Integrating multiple anthropometric variables significantly enhances predictive accuracy and offers a more comprehensive assessment of individual metabolic risk.

The results further indicate that machine learning models provide clear advantages over traditional statistical methods. Their ability to capture nonlinear relationships, interactions among variables, and complex patterns within anthropometric data allows for more precise risk stratification. Models such as random forests, gradient boosting algorithms, and neural networks consistently achieved higher discriminatory performance and demonstrated superior sensitivity in identifying individuals at elevated risk. However, the limited interpretability of some machine learning models and the need for larger, high-quality datasets highlight important considerations for their implementation in routine clinical settings.

Despite these challenges, the collective evidence affirms the potential of anthropometry-based machine learning systems to transform early diabetes prediction. As healthcare systems increasingly prioritize preventive strategies and personalized interventions, robust and scalable prediction models grounded in simple anthropometric measurements can facilitate widespread screening, guide targeted prevention programs, and reduce the long-term burden of diabetes-related complications. Future research should focus on developing hybrid models that balance interpretability and predictive power, validating algorithms across diverse populations, and integrating anthropometric prediction tools into digital health platforms to support real-time clinical decision-making. Through such efforts, anthropometric indicators can serve not only as diagnostic markers but as foundational components of comprehensive, data-driven diabetes prevention strategies.

References:

1. IDF Diabetes Atlas. (2021). *International Diabetes Federation*, 10th edition. Brussels, Belgium.
2. WHO. (2011). *Waist Circumference and Waist-Hip Ratio: Report of a WHO Expert Consultation.* Geneva: World Health Organization.

3. Hruby, A., & Hu, F. B. (2015). The epidemiology of obesity: A big picture. *Pharmacoeconomics, 33*(7), 673–689.

4. Després, J. P. (2012). Body fat distribution and risk of cardiovascular disease: An update. *Circulation, 126*(10), 1301–1313.

5. Kahn, S. E., Hull, R. L., & Utzschneider, K. M. (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature, 444*, 840–846.

6. American Diabetes Association. (2023). Classification and diagnosis of diabetes. *Diabetes Care, 46*(Suppl 1), S19–S40.

7. Wilson, P. W., Meigs, J. B., Sullivan, L., Fox, C. S., Nathan, D. M., & D’Agostino, R. B. (2007). Prediction of incident diabetes mellitus in middle-aged adults. *Archives of Internal Medicine, 167*(10), 1068–1074.

8. Schwarz, P. E., Li, J., Lindström, J., & Tuomilehto, J. (2009). Tools for predicting the risk of type 2 diabetes in daily practice. *Hormone and Metabolic Research, 41*(2), 86–97.

9. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal, 15*, 104–116.

10. Ting, D. S., Cheung, C. Y., Lim, G., et al. (2017). Deep learning in healthcare: Review, opportunities, and challenges. *Nature Medicine, 25*, 71–76.

11. Ashwell, M., & Gibson, S. (2016). Waist-to-height ratio as an indicator of early health risk: Simpler and more predictive than BMI. *BMJ Open, 6*(3), e010159.

12. Huxley, R., Mendis, S., Zheleznyakov, E., Reddy, S., & Chan, J. (2010). Body mass index, waist circumference, and waist–hip ratio as predictors of cardiovascular risk—a review. *European Journal of Clinical Nutrition, 64*(1), 16–22.

13. Browning, L. M., Hsieh, S. D., & Ashwell, M. (2010). A systematic review of waist-to-height ratio as a screening tool for the prediction of cardiovascular disease and diabetes. *Nutrition Research Reviews, 23*(2), 247–269.

14. Kodama, S., Horikawa, C., Fujihara, K., et al. (2012). Comparisons of the strength of anthropometric measures in predicting cardiovascular disease and diabetes. *Metabolism, 61*(5), 653–660.

15. Nyamdorj, R., Qiao, Q., Söderberg, S., et al. (2008). BMI compared with central obesity indicators in predicting diabetes risk. *European Journal of Clinical Nutrition, 62*, 139–147.

16. Rahman, S. A., Smith, D., & Bocquet-Appel, J. (2013). Application of machine learning methods for diabetes prediction using anthropometric data. *Journal of Biomedical Informatics, 46*(6), 1126–1132.

17. Alghamdi, M., Al-Mallah, M., Keteyian, S., et al. (2017). Predicting diabetes using supervised machine learning: A comparative analysis. *Journal of Diabetes Science and Technology, 11*(1), 80–86.

18. Choi, S. B., Kim, W. J., & Park, H. S. (2019). Machine learning-based prediction models for diabetes using anthropometric and health behavior data. *Scientific Reports, 9*, Article 16956.

19. Piwek, L., Ellis, D., Andrews, S., & Joinson, A. (2016). The rise of wearable technology in health care. *PLOS Medicine, 13*(2), e1001953.

20. WHO. (2020). *Global report on diabetes.* Geneva: World Health Organization.