

Explainable Machine Learning Framework to Prioritize Predictors of Type 2 Diabetes Risk Among Adult Males Using Lifestyle and Clinical Indicators

Siva Nanthini Shanmugam

INJE UNIVERSITY

Uchenna Esther Okpete

INJE UNIVERSITY

Md Ariful Islam Mozumder

INJE UNIVERSITY

Hee-Cheol Kim

heeki@inje.ac.kr

INJE UNIVERSITY

Article

Keywords: Diabetes Mellitus, Type 2, Machine Learning, Body Mass Index, Waist-Hip Ratio, Risk Assessment

Posted Date: September 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10940107/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

The increasing prevalent of Type 2 diabetes mellitus (T2DM) in adult males underscores the need for a risk-prediction approach that can be accurately predicted and understood. This study developed an explainable machine learning framework to compare the predictive contributions of body mass index (BMI) and waist-to-hip ratio (WHR) among adult males. Data from 2,753 adult males in the National Health and Nutrition Examination Survey 2017–2018 were analyzed using BMI-only, WHR-only, and BMI + WHR configurations. Classical machine learning and deep learning models were evaluated, and SHapley Additive exPlanations (SHAP) were used for model interpretation. Wide & Deep model with the BMI configuration achieved a test ROC-AUC of 0.9035, sensitivity of 74.74%, specificity of 84.65%, precision of 50.35%, and F1-score of 60.17% at a prespecified threshold of 0.20. Age, fasting glucose, and family history of diabetes are the top predictors based on SHAP results. When sensitivity analysis was performed by excluding fasting glucose, the Linear SVM BMI + WHR model achieved test ROC-AUC of 0.8961, with WHR showing greater significance than BMI. In independent validation, MLP - Multilayer Perceptron (BMI) model achieved the ROC-AUC was 0.8418. Explainable machine learning and maintained good discrimination after fasting-glucose exclusion and showed reproducible performance in an independent cohort.

Introduction

The identification of high-risk individuals is important to prevent the onset of type 2 diabetes mellitus (T2DM). The main challenge is that practical risk models often rely on limited predictors and fail to consider important interactions between anthropometry, lifestyle behavior, and clinical indicator [1]. The study by Lee et al. [2] showed that genetics, behavioral and socioeconomic factors could influence the occurrence of T2DM. For example, comparison of overall adiposity which is commonly assessed using body mass index (BMI) and with central adiposity, commonly assessed using waist-to-hip ratio (WHR), may provide clinically meaningful information. In a study by Sadeghi et al. [3], it was revealed that central adiposity indices such as waist-to-height ratio (WHtR) and waist to hip ratio (WHR) were more accurate in detecting diabetes amongst male compared to BMI.

In addition to excess adiposity, risk of diabetes has been consistently shown to be related to other clinical, demographic, and lifestyle factors. Studies by Jayedi [4], Tasin [5] and Rafie [6] have identified age, family history, physical inactivity, diet, lipid abnormalities, blood pressure, sleep behavior, educational attainment, and socioeconomic status as relevant factors. The growing body of evidence indicates that diabetes risk cannot be adequately quantified by one biomarker or anthropometric system of measurement alone. Prior studies supporting the comparison of anthropometric indicators and the use of explainable machine learning approaches are summarized in Supplementary Table S1. Sun [7] and Kiran [8] further indicate that diabetes mellitus emerges from the interaction between metabolic, behavioral, and social determinants stemming from complex, nonlinear interactions.

Socioeconomic and educational factors may further influence diabetes risk through differences in health literacy, occupational activity, diet quality, and access to preventive care [8]. According to Jayedi [4] and Kobayashi [9] lifestyle behaviors, including physical inactivity, sedentary behavior, unhealthy diet, excessive alcohol consumption, and sleep disruption, may interact with obesity and genetic susceptibility to increase metabolic risk [3, 4]. Rafie [6] found that family history also remains an important non-glycemic predictor, reflecting both genetic susceptibility and shared environmental exposures.

Building on the comparison between overall and central adiposity, BMI remains the most widely used anthropometric measure for diabetes screening and surveillance owing to its simplicity [9, 10]. However, it does not measure fat

distribution and visceral adiposity, which are closely associated to insulin resistance and cardiometabolic dysfunction [11]. WHR and WHtR better represent the accumulation of abdominal fat and have a higher correlation with the risk of diabetes in several populations, particularly in men [12, 13]. The relative importance of BMI and WHR for clinically relevant health outcomes is still under debate [14] and most of the current studies are based on traditional epidemiological analyses and not on a predictive model that includes BMI and these alternative measures [15].

A parallel line of work by Tassin [5] and Rafie [6] as shown in Supplementary Table S2 suggests that machine learning has become increasingly popular in diabetes risk prediction because of its capacity to model nonlinear relationships and high-dimensional interactions among clinical, behavioral, and socioeconomic variables. Despite this advantage, many ML-based diabetes models operate as black boxes, which limits their clinical acceptability and interpretability. To address this, a recent review and application study by Kiran [8] and Li [16] advocate explainable artificial intelligence (XAI) methods, such as SHAP (SHapley Additive exPlanations), to quantify individual feature contributions and support transparent, clinically defensible decision-making.

Prior research has explored applying machine learning techniques for diabetes risk classification with ensemble models, imbalance handling approaches, as well as explainability methods. Li et al [17] outline an XGBoost-based model with a class-imbalance handling and SHAP analysis for enhancing both predictive performance and interpretability of the model. Ensemble approaches using various techniques, such as stacking and boosting-based methods, have also been explored for chronic diseases risk prediction, where early detection of high-risk patients is clinically significant [18–20]. It's because given class imbalance is common for diabetes datasets, the aforementioned approaches might better identify under-represented high-risk cases. SHAP, have become widely used for interpreting complex machine learning models applied to diabetes prediction, including quantification of contributions from individual predictors, though it is still comparatively less applied in the field of explainable diabetes risk classification [21–23]. Of the various algorithms commonly evaluated, tree-based ensembles, especially the gradient boosting ones, have demonstrated strong performance due to their capability of capturing nonlinear relationships as well as interactions between heterogeneous predictors [24–26].

Despite these improvements, a lack of clarity prevails about the relative predictive value of various anthropometric indicators in different populations [27]. While BMI has been the commonly used anthropometric marker for overall adiposity, central adiposity measures such as WHR may offer additional insights regarding diabetes risk, especially in males [28, 29]. Most of the evidence comparing the performance of these anthropometric measures has primarily revolved around assessing their epidemiological relations [30] or their ability to discriminate diabetes diagnostic status [31], without assessing these measures (BMI and WHR) within a single consistent machine learning framework that also uses lifestyle, socioeconomic and clinical features to predict diabetes risk. Further methodological gap concerns the lack of interpretability of many machine learning models proposed for diabetes risk prediction [4, 6]. Consequently, an explainable modeling framework is needed to investigate the predictive information provided by BMI, and WHR, each individually, or combined, for T2DM risk in adult males and at the same time to establish the relative contributions of other important risk factors for type 2 diabetes. The principal contribution of this study is the integration of BMI-only, WHR-only, and combined BMI + WHR configurations within a unified leakage-safe, explainable machine learning framework for adult males. As part of the framework, fasting-glucose exclusion sensitivity analysis and independent validation with a separate cohort are included, allowing both non-glycemic predictors and anthropometric contributions to be assessed.

The present study aims to address these gaps by establishing an explainable machine learning model for comparing BMI and WHR in adult males based on lifestyle, socioeconomic status, anthropometric, and clinical metrics. Also, a

supplementary sensitivity analysis was conducted excluding fasting glucose to test the robustness of the framework for pre-laboratory screening. The first and main goal of this study is to compare the contribution of BMI and WHR in the prediction of T2DM risk in adult males for three different configurations of features (BMI only, WHR only and combined BMI + WHR) as depicted in Supplementary Table S3 online. Secondly, this study aims to evaluate predictive performance across multiple machine learning models using a unified pipeline for preprocessing and evaluating them. A third objective is to quantify and explain the relative contribution of diabetes-related risk factors to disease prediction using SHAP model.

Methods

Dataset Description and Study Population

The core dataset used in this study is publicly available anonymized data, sourced from the 2017–2018 cycle of the National Health and Nutrition Examination Survey (NHANES). The aforementioned National Health and Nutrition Examination Survey (NHANES) [32] led and implemented by the National Center for Health Statistics (NCHS), which is subordinate to the United States Centers for Disease Control and Prevention (CDC). It is publicly available and nationally representative. This dataset is available to everyone and does not pose any concerns of personal privacy because all personal identifying information that could be traced back to a specific individual has been removed beforehand before public release. The data used in this study were obtained from selected components collected during the NHANES 2017–2018 cycle relevant to the demographic, anthropometric, clinical, laboratory, and lifestyle variables included in the analysis. Its sampling design is reasonable, and the relevant results derived from it can represent the overall situation across the United States, rather than being limited to the conditions of a small specific area. The NHANES survey protocols were reviewed and approved by the NCHS Research Ethics Review Board, and all NHANES adult participants gave written informed consent according to the NHANES data-collection procedures. No additional data were collected for this study: instead, the study involved a secondary analysis of publicly available, de-identified NHANES data. No personally identifying information was obtained, participants were not contacted for this secondary analysis and did not involve direct participant recruitment.

The present analysis in this study was restricted to male participants aged 18 years or older. The study exclusively targeted a male population to specifically assess the predictive value of overall and central adiposity indicators, notably body mass index (BMI) and waist-to-hip ratio (WHR), in relation to the risk of type 2 diabetes mellitus (T2DM) among adult males. Conducting a male-specific analysis aimed to minimize sex-related variability in body fat distribution and anthropometric risk patterns, thereby facilitating a concentrated evaluation of diabetes risk predictors within the male demographic. After excluding participants with missing diabetes outcome information and applying predefined eligibility criteria, the final development cohort comprised 2,753 participants as shown in Fig. 1. Demographic characteristics, anthropometric measurements, blood pressure, lipid profiles, lifestyle factors, smoking status, alcohol consumption, sleep duration, education level, family history of diabetes, hypertension history, and fasting plasma glucose were included as candidate predictors. An independent external diabetes health cohort was subsequently used to evaluate the generalizability of the final selected model.

Based on Sun et al. [7], and Kobayashi et al. [9], it is known that diabetes risk profiles differ by sex, so analyses were restricted to adult males aged over 18 years. Participants with unknown diabetes outcome status were excluded from the analysis before model development [33]. Quality-control procedures were performed to identify physiologically implausible anthropometric and clinical measurements. Implausible values were treated during data quality assessment to improve the reliability of the subsequent analyses. According to Jayedi et al. [4], measurement

plausibility should be ensured for anthropometric-based diabetes prediction. Thus, an inspection of validity and missingness of continuous variables was performed before modeling was done.

As per Kobayashi et al. [9], direct diagnostic biomarkers such as HbA1c, postprandial glucose, and insulin level were excluded from the predictive set to avoid diagnostic leakage. Compared to other variables, fasting plasma glucose was retained as a numeric predictor and a data consistency variable [34]. The resulting dataset preserved the predefined predictor variables while ensuring data quality and outcome consistency for subsequent model development [35].

Outcome Definition

Based on the dataset label (diabetes), the primary outcome was binary diagnosis of type 2 diabetes mellitus (T2DM). As per Park MJ et al. [36], and Rafie et al. [6], this study followed practices in anthropometric and machine-learning-based diabetes risk modeling by treating the outcome as a classification target and not reconstructing biochemical values. Then to prevent leakage and artificially inflated performance, biomarkers that directly encode diabetes status were excluded from predictor space. This is in accordance with recommendations for risk-prediction rather than diagnostic replication by Kobayashi et al. [9]. Consequently, the final modeling is composed of lifestyle factors, sociodemographic factors, lipid factors, and anthropometric factors, including BMI and WHR, to build a screening-compatible setup. Consequently, the outcome corresponded to the diabetes status recorded in the NHANES questionnaire after applying the predefined data consistency and quality-control procedures.

Predictor Variables

Candidate predictors were selected based on clinical relevance and evidence from previous diabetes prediction studies. Exploratory data analysis was performed to summarize variable distributions and assess data quality before model development. The predictor variables comprised demographic characteristics (age and sociodemographic factors), lifestyle and behavioral indicators (physical activity, smoking, alcohol consumption, and sleep duration), anthropometric measurements (BMI and WHR), clinical history, cardiovascular risk factors (blood pressure history and lipid-related variables), and fasting plasma glucose. Three predefined feature configurations (BMI-only, WHR-only, and BMI + WHR) were evaluated while all remaining predictor variables were kept identical across configurations to isolate the predictive contribution of obesity-related anthropometric measures. According to Jayedi et al. [4] & Park MJ et al. [33], models integrating obesity indicators with behavioral and clinical risk factors improve diabetes risk prediction.

Data Preprocessing

All preprocessing procedures were implemented using a leakage-safe machine learning pipeline as shown in Fig. 2. The development dataset was first divided into training, validation, and independent NHANES test sets before any preprocessing operations were performed. K-nearest neighbors (KNN) imputation was used to imputed numerical values that are missing from the database which was fitted only to the training data and the imputed values were then applied to the validation and test data sets. A pipeline of ColumnTransformer was used to standardize numerical variables using StandardScaler and one-hot encoding was used to transform categorical variables. Identical preprocessing procedures were applied across all feature configurations to ensure fair model comparison while preventing information leakage between data partitions.

The independent validation dataset underwent an independent preprocessing procedure before model evaluation. Numerical variables were standardized, and categorical variables were one-hot encoded using the same feature definitions as the development dataset [34]. Records with missing values in key anthropometric variables (BMI or WHR), contradictory fasting glucose–diabetes labels, or missing outcome information were excluded during data cleaning. Consequently, the independent validation data set constituted a complete-case cohort, and no additional imputation was required before evaluation [24].

Feature Configuration Strategy

A mutually exclusive set of feature configurations were compared under identical preprocessing and evaluation conditions in order to isolate and compare the predictive contribution of BMI versus WHR. A configuration-based comparison follows the logic that obesity indicators differ from each other and can carry different information about diabetes risk [11, 34].

Machine Learning Models

Each configuration was trained with six different classical machine learning models including Logistic Regression, Linear SVM (calibrated), Random Forest, Balanced Random Forest, XGBoost, and CatBoost. A wide range of approaches was chosen to cover linear baselines, margin-based classifiers, bagging ensembles, imbalance-aware ensembles, and gradient-boosted decision trees [35]. Based on tabular clinical and lifestyle datasets, these models frequently achieve strong discrimination [36] in diabetes prediction.

In addition to the well-known machine learning models, three deep learning benchmark architectures were evaluated: multilayer perceptron (MLP), a Wide & Deep neural network, and a residual neural network, or ResNet. This set of architectures were considered as additional benchmarks to ascertain whether increased neural network complexity led to a better prediction performance when working on the structured tabular dataset. Each deep learning architecture was trained using five different random seeds, and the mean and standard deviation of the validation ROC-AUC were reported to assess model stability during model selection.

Model Training and Evaluation

The development dataset was divided into training (64%), validation (16%), and independent test (20%) subsets using stratified sampling to preserve the prevalence of diabetes across all partitions. The model selection was carried out separately according to the type of model. As for classical machine-learning models, hyperparameter tuning and candidate selection were based on mean ROC-AUCs computed from three-fold stratified cross-validation performed exclusively on the training data. Candidates for deep-learning models were selected based on their performance on held-out validation set. As a result of this separation, the independent test set was not accessed while the model and hyperparameter selection were being completed.

Upon selecting the final candidate, the model configuration was locked before evaluating the test-set. We then refitted the selected model using the combined training and validation data. In the final deep-learning model, the number of training epochs was fixed at 17, which was based on the median optimal epoch determined through the development process. Thus, the model can train without using test-set information because it uses the entire non-test development cohort. The independent NHANES test set remained completely untouched throughout model development as stated by Kobayashi et al. [9], and Tasin et al. [5]. Once the best-performing model per feature

configuration was selected, the final model was retrained using both the training and validation datasets and tested once with the independent NHANES test set [37–39].

To better reflect a screening-focused clinical scenario, a prespecified fixed decision threshold of 0.20 favoring sensitivity while keeping specificity reasonably high was chosen for computing threshold-dependent performance measures, such as accuracy, precision, recall, F1-score, specificity and negative predictive value. The threshold was fixed before evaluation on the independent NHANES test set and was not optimized using test-set performance. Discrimination performance was assessed using the area under the receiver operating characteristic curve (ROC-AUC).

Finally, the complete modeling framework was independently evaluated on the Diabetes Health Indicators dataset which initially comprised approximately 100,000 participants. After restriction to adult males aged ≥ 18 years and application of the predefined data-screening and label-cleaning criteria, 22,012 participants constituted the final analytical cohort. The cohort was stratified into training ($n = 14,087$; 64%), validation ($n = 3,522$; 16%), and independent hold-out test ($n = 4,403$; 20%) sets. In this cohort, the same model-development protocol was followed: classical models were selected by three-fold stratified cross-validation within the training data, whereas deep learning models were selected by using the validation data. Independent hold-out tests remain separate from model selection and were used only for final evaluation. The same preprocessing strategy, feature configurations, model architectures, hyperparameter settings, and prespecified decision threshold used in the NHANES framework were applied for Diabetes Health Indicators dataset as an independent validation, is illustrated in Supplementary Figures S5 and S6.

Class Imbalance Handling

Class imbalance was quantified within the NHANES development cohort by calculating the ratio of diabetic to non-diabetic participants [40] before model development as reported in the studies by Park MJ et al. [33], and Tasin et al. [5]. Instead of synthetic oversampling, model-based strategies were used to handle imbalances. Specifically, class-weighting strategies were applied to Logistic Regression, Linear SVM, and Random Forest models to increase the contribution of the minority class during training. For CatBoost, the built-in `auto_class_weights="Balanced"` option was employed to automatically adjust class weights during training [17, 41]. `Scale_pos_weight` is derived from the training-set class ratio of XGBoost models, a common technique in gradient-boosting pipelines for imbalanced clinical outcomes [36]. For the final experiments, no synthetic oversampling was applied (namely, SMOTE or ADASYN) so that performance could reflect the actual distribution of data. This still benefits from imbalance-aware learning rules, in accordance with Tasin et al. [5] & Rafie et al. [6] recommendations for transparent evaluation of chronic disease classifiers.

Explainable Artificial Intelligence Analysis

SHapley Additive exPlanations (SHAP) were employed to improve the interpretability of the selected prediction models beyond conventional predictive performance metrics. SHAP analyses were performed for the final selected models and feature configurations to quantify the contribution of individual predictors to model predictions at both global and local levels. For the BMI configuration, SHAP analysis was applied to the final Wide & Deep model, whereas for the combined BMI + WHR configuration, SHAP analysis was applied to the final Linear SVM model. SHAP values were calculated using explainers appropriate for the respective model architectures after applying the corresponding fitted preprocessing procedures. Global feature importance was evaluated using SHAP summary plots and mean absolute SHAP values to rank predictors according to their overall contributions to model

predictions. Particular attention was given to the relative contributions of the anthropometric measures, including BMI and WHR, in relation to other demographic, lifestyle, and clinical predictors. Local SHAP values were additionally examined to support the interpretation of individual predictions and to illustrate how specific risk factors contributed to predicted diabetes status. This explainability analysis was used to provide transparent interpretation of the final predictive models and to facilitate comparison of the roles of overall and central adiposity within the diabetes prediction.

Results

Baseline Characteristics of the Study Population

After implementing the selection criteria and conducting the data screening process, 2,753 adult males (aged 18 years or older) were included in the study sample for the NHANES 2017 – 2018 development cohort. Demographic features, lifestyle and behavioral factors, anthropometric measurements (BMI and WHR), cardio metric parameters including blood pressure and lipid profiles, fasting plasma glucose, family history of diabetes and history of hypertension were included as predictor variables as shown in Supplementary Table S4 online. Out of eligible population, 2276 (82.67%) were non-diabetic and 477(17.33%) were diabetic. The development group was then split into a training dataset (n = 1,761), a validation dataset (n = 441), and an independent testing dataset (n = 551), with the class distribution being the same in each of the three data partitions, 64%–16%–20%.

Model Performance Across Feature Configurations (BMI vs WHR vs BMI+WHR)

The model performance was assessed with receiver operating characteristic area under the curve (ROC–AUC) and classification metrics that depend on the threshold (accuracy, precision, recall, F1-score, specificity and negative predictive value (NPV). Candidate models were selected according to the prespecified model-selection criteria described in the model training and evaluation section, while the untouched test set was reserved exclusively for the final evaluation to prevent selection bias. Across the three predefined feature configurations (BMI, WHR, and BMI+WHR), all classical machine learning and deep learning models demonstrated strong discriminative performance, with validation ROC–AUC values generally ranging from approximately 0.88 to 0.91. Among the classical machine learning algorithms, logistic regression and linear support vector machine consistently achieved the highest mean cross-validation performance across the three feature configurations, whereas tree-based ensemble models exhibited comparable but slightly lower discrimination. Among the deep learning architectures, the WideDeep model achieved the highest validation ROC-AUC performance using the BMI feature configuration and was therefore selected as the final candidate according to the prespecified model-selection. Overall, predictive performance was similar across the BMI, WHR, and BMI+WHR configurations. The detailed performance metrics for all evaluated models and feature configurations are presented (Tables 1-3, and Supplementary Table S5 online),

Selected operating point and threshold-based behavior

To facilitate clinically meaningful interpretation of the prediction results, the final selected WideDeep model using the BMI feature configuration was evaluated at the prespecified probability threshold of 0.20. The model was able to predict 386 true negative and 71 true positives, 70 false positives and 24 false negative on the hold-out test set i.e for sensitivity or recall of 74.74%, specificity of 84.65%, precision of 50.35% and F1 value of 60.17 % with an overall test

ROC–AUC value of 0.9035. The confusion matrix is shown in Supplementary Figure S1, while detailed threshold-dependent performance metrics for best classical models and feature configurations are summarized in Table 3 and Supplementary Table S5 online.

Explainability Results

To improve transparency and clinical interpretability, we performed SHAP analysis (SHapley Additive exPlanations) on the final selected WideDeep model using the BMI feature configuration. The global SHAP analysis revealed age as the most prominent risk factor for developing the disease with the mean absolute SHAP value of 0.0809. The complete ranking of feature importance is presented in Table 4, while the corresponding global SHAP visualizations are shown in Figure 3 and Supplementary Figure S2.

The SHAP summary plot further revealed the direction of feature effects on model predictions. Higher values of age, fasting plasma glucose, BMI, heart rate, family history of diabetes, and hypertension history generally produced positive SHAP values, thereby increasing the predicted probability of diabetes. Conversely, lower values of these variables tended to decrease the predicted risk. Similarly, higher levels of LDL-cholesterol tended to have a positive association with the model's output, while increased physical activity, extended sleep duration, and superior socioeconomic status, as indicated by the poverty-income ratio, tended to have a lower impact on positive diabetes predictions.

The local SHAP explanation further demonstrates the interpretability of the proposed model at the individual level. For the pictured individual, family history of diabetes, total cholesterol, BMI, and systolic blood pressure were positive factors in increasing the predicted probability of diabetes, but diastolic blood pressure, fasting plasma glucose, and LDL cholesterol, physical activity, heart rate, absence of hypertension history, and a few other variables were negative factors in decreasing the model output.

Sensitivity Analysis Excluding Fasting Glucose

To assess the predictive performance of the proposed framework in the absence of a glycemic biomarker, an additional sensitivity analysis was conducted after excluding fasting plasma glucose from all predictor configurations. The same preprocessing procedures, BMI, WHR, and BOTH anthropometric configurations, classical and deep-learning architectures, class-weighting strategy, and leakage-safe evaluation framework used in the primary analysis were retained. The classification threshold was fixed at 0.20, consistent with the screening-oriented operating point prespecified for the primary analysis, rather than re-optimized within the sensitivity analysis. The present analysis included the same 2,753 NHANES adult male participants (split into training, validation and independent NHANES test sets of 1,761, 441, and 551 participants, respectively). The same prespecified model selection procedure described in the model training and evaluation section was applied, while the test set remained unused during model and feature-configuration selection.

Following exclusion of fasting glucose, the Linear Support Vector Machine under the BOTH configuration (BMI + WHR) achieved the highest selection performance, with a mean training cross-validation ROC-AUC of 0.8658 ± 0.0103 and a validation ROC-AUC of 0.8807. It was therefore locked as the final sensitivity-analysis candidate based on the prespecified model selection procedure. The final model achieved a test ROC-AUC of 0.8961 (95%CI: 0.8623-0.9232), accuracy of 82.03% (95%CI: 78.77-85.30%), sensitivity of 75.79% (95%CI: 67.01-84.00%), specificity of 83.33% (95%CI: 79.82-86.73%), precision of 48.65% (95%CI: 40.49-56.79%), and an F1-score of 59.26% (95%CI: 51.03-

66.17%). The negative predictive value was 94.29%. At the fixed threshold of 0.20, the confusion matrix comprised 380 true negatives, 76 false positives, 23 false negatives, and 72 true positives. The findings in this study suggest that the framework was still able to retain a high discriminability even in the absence of fasting glucose measurements. However, its test ROC-AUC was slightly lower than that of the original WideDeep-BMI framework with fasting glucose.

SHAP analysis of the final LinearSVM-BOTH models as shown in Figure 4 and Supplementary Figure S3 demonstrated a redistribution of feature importance after fasting glucose was removed. The most leading global factor was age, with a mean of the absolute SHAP value of 0.0809, followed by family history of diabetes 0.0581, hypertension history 0.0386, LDL-cholesterol 0.0296, heart rate 0.0272, waist to hip ratio 0.0210, diastolic blood pressure 0.0188, HDL-cholesterol 0.0170 and 0.01354 for Non-Hispanic White ethnicity and BMI 0.0125. The less significantly contributed factors are physical activity, sleep duration, alcohol consumption, education, smoking history, systolic blood pressure, triglycerides, poverty-income ratio, and total cholesterol. On this analysis, the SHAP values from WHR showed greater global feature importance than those from BMI as indicated by its higher mean absolute SHAP value (0.0210 vs. 0.0125) within the glucose exclusion model. This comparison reflects relative global feature importance within the selected model and should not be interpreted as evidence of statistical significance or causality.

Among the deep-learning models, the MLP based on the BMI architecture achieves the highest average validation ROC-AUC value 0.8573 ± 0.0029 , followed closely by that of the WideDeep-BMI, with 0.8562 ± 0.0044 . According to the prespecified model selection procedure, these validation results were lower than the selection performance of LinearSVM-BOTH, indicating that the more complex neural architecture did not provide a performance advantage during candidate selection in the glucose-exclusion setting. Detailed candidate-model results are provided (Supplementary Tables S6-S8).

Independent Validation on the Diabetes Health Indicators Dataset

In the independent Diabetes Health Indicators cohort, the selected MLP model (BMI configuration) evaluated on the 20% independent hold-out test set ($n=4,403$) yielded a ROC-AUC of 0.8418, an accuracy of 0.60, sensitivity of ~ 0.95 , specificity of 0.21, a F1-score of 0.71. With a pre-defined screening threshold of 0.20, the confusion matrix comprised 458 true negatives, 1,656 false positives, 92 false negatives, and 2,197 true positives. In the glucose-inclusive analysis, SHAP analysis identified fasting glucose as the predominant predictor, followed by family history of diabetes and age, with physical activity, screen time, bmi, total cholesterol, hdl cholesterol, ldl cholesterol, diet score and other clinical and lifestyle factors contributing to prediction. A glucose-exclusion sensitivity analysis was also performed in the independent validation cohort. Under this analysis, the selected LinearSVM model using the BMI configuration achieved a test ROC-AUC of 0.7477. The sensitivity, specificity, precision, and negative predictive value at a selected operating point, as obtained by the model, were 97.90%, 6.81%, 53.21%, and 75%, respectively, along with 144 true negatives, 1,970 false positives, 48 false negatives, and 2,241 true positives. The feature that contributed most to the predictions was family history of diabetes, as found by SHAP followed by age, physical activity, BMI, diet score, screen time, HDL cholesterol, triglycerides, smoking status, LDL cholesterol and heart rate. Detailed validation results, including model comparisons, threshold-based performance metrics, SHAP analyses, and glucose-exclusion sensitivity analyses, are provided in Supplementary Tables S9-S14 online.

Discussion

This study developed and evaluated an explainable machine learning framework for diabetes risk stratification in adult males using demographic, lifestyle, anthropometric, and clinical characteristics derived from the NHANES

2017–2018 cohort. Three anthropometric feature configurations (BMI-only, WHR-only, and BMI + WHR) were systematically compared under an identical leakage-safe machine learning pipeline with prespecified selection procedure. According to that prespecified procedure, the Wide & Deep architecture with the BMI-only feature configuration yielded high discrimination rates on the independent NHANES test set coupled with a clinically meaningful balance of sensitivity vs specificity at the predefined screening cutoff threshold of 0.20. This operating point does not represent a universally optimal decision threshold; instead, it offers a clinically interpretable balance between sensitivity and specificity when evaluated in conjunction with the threshold-independent ROC–AUC and the explainability analyses [7, 37]. Additionally, the proposed framework underwent independent validation using the Diabetes Health Indicators dataset. This validation demonstrated that the proposed modeling strategy maintained predictive performance in an independent population, despite variations in dataset composition and participant characteristics.

Two primary findings were identified in this study. Firstly, the predictive performance of all three anthropometric configurations was similar, indicating that both BMI and WHR are valuable for predicting diabetes risk. Nevertheless, both the WHR only and BMI + WHR configurations exhibited only modest improvements on the BMI only configuration suggesting that BMI and WHR capture partially overlapping information relevant for diabetes risk. Second, SHAP analysis showed that age was the single most important diabetes risk factor followed by fasting glucose level, family history of diabetes, history of hypertension, diastolic blood pressure, heart rate, LDL cholesterol, ethnicity, BMI, HDL cholesterol. This suggests that glycemic status and familial predisposition are more influential predictors than anthropometric measurements alone, while BMI and WHR provide additional information relevant to better individualized risk assessment. At a general overview, the SHAP patterns suggest that diabetes risk in adult males involves a complex interplay of demographic, metabolic, cardiovascular, anthropometric, behavioral, and socioeconomic factors rather than by any single determinant.

To evaluate the applicability of the proposed framework in settings where laboratory biomarkers are unavailable, an additional sensitivity analysis was performed after excluding fasting glucose from model development. Although discrimination decreased relative to the primary glucose-inclusive analysis, the NHANES glucose-exclusion analysis identified age as the leading SHAP-ranked predictor, followed by family history of diabetes, hypertension history, LDL cholesterol, heart rate, WHR, diastolic blood pressure, HDL cholesterol, ethnicity and BMI. A similar reduction in discrimination was observed in the independent Diabetes Health Indicators cohort; however, the relative importance of the predictors differed. In this cohort, the leading SHAP ranked factors were family history of diabetes followed by age, and physical activity, along with BMI also being an important predictor. Although there were some discrepancies in the feature importance rankings across the two cohorts, age, and family history of diabetes were among the top non-glycemic factors in both analyses. The results of this study indicate that demographics, familial, anthropometric, cardiovascular and lifestyle factors provide information for prediction of diabetes risk in the absence of fasting glucose. However, the loss in discrimination after removing fasting glucose suggests that such models may serve as a complementary tool for preliminary risk stratification, rather than replace confirmatory laboratory testing or clinical diagnosis.

A wide variety of epidemiologic investigations have shown that general and central obesity are linked with diabetes risk type 2 [4]. In fact, multiple meta-analyses reported on robust dose-response relationships between diabetes risk and BMI, waist circumference, or waist-hip ratio, which further support the hypothesis that fat distribution in the abdominal region is a more pertinent factor than the overall body mass in determining metabolic dysfunction [42–44]. Similarly, multiple clinical prediction studies have reported that either the WHR or the waist to height ratio may be more pertinent to the prediction of impaired glucose regulation than the BMI, particularly in male populations [11].

The findings from the present study are widely aligned with this literature, however, offering a more nuanced interpretation in a broad machine learning context. BMI and WHR were both positively correlated with diabetes risk, and both showed a meaningful predicative content [7]. Nevertheless, in contrast with investigations mainly based on anthropometric values, only slightly differences between BMI only and WHR only models were found, and including both indices provides only limited improvement in additional predictive discrimination. These findings point to the fact that when anthropometric variables are assessed in concert with fasting glucose, family history, cardiovascular measurements, and behavioral data, the relative importance of a specific obesity index becomes less pronounced [4].

The observed differences to past reports may be explained by methodological choices rather than any biological inconsistencies. Earlier studies frequently evaluated anthropometric indices either in isolation or within conventional statistical models containing relatively few clinical variables [8, 45]. In contrast, the current framework integrated demographic, behavioral, anthropometric and laboratory measurements into an explainable machine learning pipeline [17]. Thus, fasting glucose and family history were found to be the dominant contributors to the prediction, while BMI and WHR played complementary rather than dominant roles [2]. These findings support previous observations that the clinical utility of obesity indices depends strongly on the broader predictor set included within the predictive model.

The sensitivity analysis further extends previous work by demonstrating that predictive information remained available even after fasting glucose was excluded [2]. Although reduced discrimination is compared with the glucose-inclusive analysis, both the NHANES and independent validation analyses showed that demographic, familial, anthropometric, behavioral, and cardiovascular characteristics retained predictive value in the absence of fasting glucose. These findings support further investigation of non-laboratory machine-learning approaches for preliminary diabetes risk stratification before confirmatory laboratory assessment.

The present findings have several practical implications for diabetes risk assessment in adult males. First, integrating demographic, lifestyle, anthropometric, cardiovascular, and laboratory variables within an explainable machine learning framework provides substantially richer clinical information than relying solely on individual anthropometric indices. The SHAP analysis demonstrated that fasting glucose, family history of diabetes, age, hypertension history, and cardiovascular risk markers contributed most strongly to individualized risk prediction, whereas BMI and WHR provided complementary but comparatively smaller contributions.

Second, the long-standing debate regarding whether BMI or WHR represents the superior anthropometric marker becomes less clinically relevant once broader clinical information is incorporated into prediction models. Although both indices remained informative, their predictive contributions were considerably smaller than those of fasting glucose and family history. These findings suggest that comprehensive clinical assessment may be considerably more beneficial, rather than identifying an individual anthropometric metric, by the standard diabetes risk assessment practice.

Third, the sensitivity analysis implies that the presented framework may be used to support diabetes screening in the community where the laboratory test is unavailable, too expensive, or delayed. Under these circumstances, the explainable machine learning, based on anthropometric indices, family history information, cardiovascular markers, and behavioral features, may be used to select candidates for whom diagnostic test will be given for diagnosis confirmation. This approach can improve early identification of high-risk individuals but support effective deployment of health resources.

Lastly, SHAP explanation is useful to improve the clinical interpretability of the proposed framework. The model does not simply return estimates probability of outcome but also explains the underlying components that make up each

of these predictions, enabling clinicians to offer tailored risk information and to focus on potentially modifiable components, i.e. obesity, hypertension, physical inactivity, and dysglycemia when providing counselling. In addition, the analysis is restricted to adult males to respond to the growing need of sex-specific risk stratification for diabetes since men and women differ risk factor profiles and disease progression. Altogether, the findings of this study are in favor of the potential use of explainable machine learning for interpretable clinical decision support system for individual diabetes risk assessment and early preventive intervention.

This study has several strengths relative to previous studies comparing BMI/WHR and diabetes machine learning frameworks. The study used a national-representative NHANES cohort comprising 2,753 adult males which enables leakage-safe model development and independent testing via stratified data partitioning. This study design adheres to current best-practices recommendations for chronic disease machine-learning pipelines where model selection, validation, and independent testing are separated to reduce information leakage [6]. Furthermore, the male-only study analysis directly responds to epidemiologic concerns that adiposity-risk associations may differ according to the sex and central obesity might be especially informative for men [6, 44]. Under the same conditions, a common preprocessing pipeline and model-specific selection procedures are employed to make a fair comparison between BMI and WHR and BMI + WHR combinations.

Class imbalance, was addressed through algorithm-specific approaches, including class weighting, Balanced Random Forest, and XGBoost's `scale_pos_` [17, 36] weight, thereby improving robustness without relying on synthetic oversampling. It is important to note that explainability is integrated and not added post-hoc, allowing the contribution of adiposity indicators to stronger clinical and familial predictors to be visualized and quantified [46, 47]. Fasting glucose served two complementary purposes: it was used during data cleaning to identify inconsistent outcome labels and was subsequently included as a clinically relevant predictor during model development. This approach improved label integrity while allowing the contribution of fasting glucose to be explicitly quantified through SHAP analysis [9]. An additional strength of this study is the independent validation of the complete modeling framework using a separate Diabetes Health Indicators cohort. Applying the same preprocessing strategy, model architectures, hyperparameter settings, prespecified model selection procedure, and decision threshold to an independent population demonstrated that the proposed methodology remained robust beyond the NHANES development cohort. This provides stronger evidence for methodological reproducibility than evaluation on a single dataset alone.

There are several limitations that need to be acknowledged. First, because the design is cross-sectional, causal inference cannot be conducted and associations represent contemporaneous rather than temporal relationships. This is consistent with limitations in previous diabetes ML studies [5, 48]. Second, several variables in NHANES, including physical activity, alcohol consumption, sleep duration, smoking history, and family history of diabetes, were obtained through standardized participant questionnaires rather than objective measurements. Consequently, these variables may be affected by recall bias and reporting bias, potentially attenuating their estimated contribution to diabetes risk prediction [49]. Third, while fasting glucose was strongly associated with diabetes diagnosis and hence dominated the SHAP-importance ranking in primary analyses, the relative importance of behavioral anthropometric predictors may have been underestimated. Although fasting glucose significantly influenced the prediction and led to a dominant position in the SHAP ranking, sensitivity analysis revealed that predictive discrimination can still be achieved after removing the fasting glucose, indicating that anthropometric, cardiovascular and familial predictors persist in informative characteristics for assessing diabetes risk. In addition, because only the prespecified LinearSVM-BOTH model was evaluated on the independent NHANES test set in the glucose-exclusion sensitivity analysis, its test-set performance should not be interpreted as evidence of superiority over alternative models that were not evaluated on the test set. Although the proposed framework showed consistent performance in an

independent validation dataset, further evaluation on prospective cohorts, healthcare systems and geographically diverse populations is needed before clinical wide implementation. A fourth limitation was the absence of direct measures of body composition, thus making it impossible to make a distinction between fat distribution and lean mass reduction. This has been shown to play a significant role in the development of diabetes risk profiles in males [9, 50]. Although the proposed framework was independently validated using the Diabetes Health Indicators dataset, both datasets originated from the United States. Consequently, additional validation across healthcare systems with different demographic, socioeconomic, and ethnic characteristics is necessary to establish broader international generalizability as per prior ML reviews from Tasin et al. [5] and Rafie et al. [6].

Future work should focus on broader independent validation across geographically diverse populations to determine whether the proposed framework and SHAP-derived feature rankings remain stable across different healthcare settings and ethnic groups. Second, these future studies should further investigate diabetes risk prediction both with and without glycemic biomarkers to better understand the contribution of anthropometric, behavioral, and cardiovascular variables during pre-diagnostic screening. Comparisons across multiple populations may further clarify the extent to which non-laboratory predictors can compensate for the absence of biochemical testing while supporting scalable community-based screening programs. As a third consideration, extending datasets to include direct body-composition metrics such as DXA-derived visceral fat or muscle mass metrics may improve mechanistic understandings of adiposity-driven risk in men. Fourth consideration is the development of deployment-oriented XAI interfaces for translating SHAP values into clinician-friendly explanations. Finally, future work should investigate prospective implementation of the proposed framework within clinical decision-support systems to evaluate calibration stability, clinician usability, workflow integration, and its influence on diabetes prevention strategies in routine clinical practice.

Conclusion

To summarize, this study established a machine learning model for T2DM risk prediction in adult males that is leakage-safe and explainable in its relative contribution of BMI and WHR. BMI and WHR were both informative in predicting T2DM risk with demographic, lifestyle and clinical characteristics, though only small differences in model performance between anthropometric configurations were found. SHAP-based analysis offered clear interpretation of the importance of factors that contributed to model predictions, while glucose-exclusion sensitivity analyses showed that non-glycemic traits maintained predictive risk information with fasting glucose missing, though with some loss of model discrimination. The overall modeling framework was reproducible in independent validation using a separate population Diabetes Health Indicators dataset. Overall, these findings support further evaluation to examine explainable machine learning methods for individualized diabetes risk stratification, with future research needed to examine the performance in longitudinal and geographically diverse populations and their potential integration into clinical decision-support systems.

Declarations

Additional information

Competing Interests:

The authors declare no competing interest.

Funding

Declaration

This research was funded by the Institute of Information & Communications Technology Planning & Evaluation (IITP) - Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government (MSIT) (IITP-2024-RS-2024-00436773).

Author Contribution

S.N.S. performed the data acquisition, analysis, and interpretation and prepared the original manuscript draft. S.N.S., U.E.O., M.A.I.M., and H.C.K. contributed to the conceptualization and study design. U.E.O., M.A.I.M., and H.C.K. contributed to manuscript review and revision. All authors reviewed and approved the final manuscript.

Acknowledgement

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) through the Innovative Human Resource Development for Local Intellectualization Program, funded by the Korea government (MSIT) (IITP-2024-RS-2024-00436773).

Data Availability

The datasets analyzed during the current study are publicly available. The National Health and Nutrition Examination Survey (NHANES) 2017–2018 dataset used for the primary analysis is available from the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC) at <https://wwwn.cdc.gov/nchs/nhanes/> and <https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017>. The Diabetes Health Indicators Dataset used for independent validation is publicly available on Kaggle at <https://www.kaggle.com/datasets/mohankrishnathalla/diabetes-health-indicators-dataset>. No new participant-level datasets were generated during this study.

References

1. Lee, H., Park, M. B. & Won, Y. J. AI machine learning-based diabetes prediction in older adults in south korea: Cross-sectional analysis. *JMIR Formative Res.* **9**, e57874 (2025).
2. Lee, H. et al. Prediction model for type 2 diabetes mellitus and its association with mortality using machine learning in three independent cohorts from south korea, japan, and the uk: a model development and validation study. *EClinicalMedicine* **80**, 103069 (2025).
3. Sadeghi, E. et al. Novel anthropometric indices for predicting type 2 diabetes mellitus. *BMC Public. Health.* **24**, 1033 (2024).
4. Jayedi, A. et al. Anthropometric and adiposity indicators and risk of type 2 diabetes: systematic review and dose-response meta-analysis of cohort studies. *BMJ* **376**, e067516 (2022).
5. Tasin, I., Nabil, T. U., Islam, S. & Khan, R. Diabetes prediction using machine learning and explainable AI techniques. *Healthc. Technol. Lett.* **10**, 1–10 (2023).

6. Rafie, Z. et al. Leveraging XGBoost and explainable AI for accurate prediction of type 2 diabetes. *BMC Public Health*. **25**, 3688 (2025).
7. Sun, Q. et al. Machine learning-based assessment of diabetes risk. *Applied Intelligence* **55** (2024).
8. Kiran, M. et al. Type 2 diabetes prediction without labs: a systems-level neural framework for risk and behavioral network reorganization. *Front. Digit. Health*. **7**, 1714545 (2025).
9. Kobayashi, G. et al. Associations between anthropometric indices as complementary predictors and incidence of type 2 diabetes; Panasonic cohort study 21. *Diabetes Res. Clin. Pract.* **217**, 111888 (2024).
10. Lin, J. C. & Tandar, C. E. Global patterns of visceral adiposity and car-diometabolic disease. *Nat. Reviews Endocrinol.* **22**, 8 (2026).
11. Sun, K. et al. Assessment of adiposity distribution and its association with diabetes and insulin resistance: a population-based study. *Diabetol. Metab. Syndr.* **11**, 51 (2019).
12. Makinen, V. P., Zhao, S., Ihanus, A., Tynkkynen, T. & Ala-Korpela, M. Epi-demiological associations between obesity, metabolism and disease risk: are body mass index and waist-hip ratio all you need? *Int. J. Obes.* **49**, 2555–2560 (2025).
13. Ma, Y. L. et al. Waist-to-height ratio is a simple and practical alternative to waist circumference to diagnose metabolic syndrome in type 2 diabetes. *Front. Nutr.* **9**, 986090 (2022).
14. Lee, T. L. et al. Evaluating the potential of waist-to-BMI ratio, a body shape index, and other anthropometric parameters in predicting cardiovascular disease mortality: evidence from NHANES III. *BMC Public Health*. **25**, 1828 (2025).
15. Li, W., Peng, Y. & Xiong, S. Diabetes risk prediction model based on improved XGBoost with genetic algorithm. Proceedings of the 2024 4th International Conference on Computational Modeling, Simulation and Data Analysis Pp. 390–396. (2024).
16. Li, W., Peng, Y. & Peng, K. Diabetes prediction model based on GA-XGBoost and stacking ensemble algorithm. *PLOS ONE*. **19**, e0311222 (2024).
17. Mahajan, P., Uddin, S., Hajati, F., Moni, M. A. & Gide, E. A comparative evaluation of machine learning ensemble approaches for disease prediction using multiple datasets. *Health Technol.* **14**, 597–613 (2024).
18. Farnoosh, R., Abnoosian, K. & Isewid, R. A. Two machine-learning hybrid models for predicting type 2 diabetes mellitus. *J. Med. Signals Sens.* **15**, 11 (2025).
19. Ge, X. et al. Application of machine learning tools: Potential and useful approach for the prediction of type 2 diabetes mellitus based on the gut microbiome profile. *Experimental Therapeutic Med.* **23**, 305 (2022).
20. Jithendra, V., Mohit, R. M. S., Madhusudhan, M., Jagadeesh, B. & Kusuma, S. Diabetes prediction using machine learning techniques. *J. Artif. Intell. Capsule Networks*. **5**, 190–206 (2023).
21. Yousif, D. et al. Gender-specific fat distribution and association with type 2 diabetes mellitus: A study of adiposity indicators (DEXA, BMI, waist-to-hip ratio) in adults from Qatar. *Diabetes Metabolic Syndrome Obes.* **17**, 1715–1724 (2024).
22. Boonpor, J. et al. Associations and predictive performance of 11 anthropometric measures with incident type 2 diabetes: A prospective cohort study from the UK biobank. *Obesity* **31**, 2648–2657 (2023).
23. Seo, B. M. & Choi, J. W. Construction of risk prediction model of type 2 diabetic kidney disease based on deep learning (diabetes metab j 2024. *Diabetes Metabolism J.* **48**, 48, 771–779 (2024).
24. Srinivasu, P. N., Ahmed, S. & Hassaballah, M. Almusallam, N. An explainable artificial intelligence software system for predicting diabetes. *Heliyon* **10**, e36112 (2024).

25. Ghomi, F., Sefidkar, R., Khaledi, E. & Jambarsang, S. Optimal cut-off points of anthropometric and body roundness indices associated with diabetes: Persian (shahedieh) cohort study. *Front. Nutr.* **11**, 1428704 (2024).
26. Arrington, C. E. et al. Genetic determinants of BMI, diet, and fitness interact to partially explain anthropometric obesity traits but not the metabolic consequences of obesity in men and women. *Int. J. Obes.* **50**, 938–946 (2026).
27. Gandhi, V. et al. Benchmarking machine learning models for obesity classification with SHAP-based interpretability. *International J. Comput. Intell. Systems* **19** (2025).
28. Lee, B. J. & Yim, M. H. Comparison of anthropometric and body composition indices in the identification of metabolic risk factors. *Sci. Rep.* **11**, 9931 (2021).
29. Maimaitijiang, E., Aihaiti, M. & Mamatjan, Y. An explainable AI framework for online diabetes risk prediction with a personalized chatbot assistant. *Electronics* **14** (2025).
30. Wardhani, K. D. K. & Akbar, M. Diabetes risk prediction using extreme gradient boosting (XGBoost). *Jurnal Online Informatika.* **7**, 244–250 (2022).
31. Afolabi, S., Ajadi, N., Jimoh, A. & Adenekan, I. Predicting diabetes using supervised machine learning algorithms on e-health records. *Inf. Health.* **2**, 9–16 (2025).
32. NCHS. National health and nutrition examination survey (NHANES) questionnaires, datasets, and related documentation. (2012).
33. Park, M. J. et al. A novel anthropometric parameter, weight-adjusted waist index represents sarcopenic obesity in newly diagnosed type 2 diabetes mellitus. *J. Obes. Metabolic Syndrome.* **32**, 130–140 (2023).
34. Rizal, Azzahra, A. F., Bandaso, R. M. S., Harefa, J. & Jingga, K. Type 2 diabetes prediction using recall-optimized XGBoost. *Procedia Comput. Sci.* **269**, 1309–1318 (2025).
35. Luo, J. et al. Association between anthropometric indices and early arterial stiffness in patients with type 2 diabetes in china: a cross-sectional study. *BMJ Open.* **15**, e097626 (2025).
36. Lin, D. N. et al. Elevated waist-to-hip ratio, as an abdominal obesity index, predicts the risk of diabetic kidney injury. *World J. Diabetes.* **16**, 101384 (2025).
37. Yin, J. M. et al. Explainable machine learning-based prediction model for diabetic nephropathy. *Journal of Diabetes Research* 8857453 (2024). (2024).
38. Ejiyi, C. J. et al. A robust predictive diagnosis model for diabetes mellitus using shapley-incorporated machine learning algorithms. *Healthcare Analytics* **3** (2023).
39. Rafsanjani, K. et al. Waist-to-hip ratio as a contributor associated with higher atherosclerotic cardiovascular disease risk assessment in patients with diabetes: a cross-sectional study. *BMC Cardiovasc. Disord.* **24**, 613 (2024).
40. Ganie, S. M., Pramanik, P. K. D., Malik, B., Mallik, M., Qin, H. & S. & An ensemble learning approach for diabetes prediction using boosting techniques. *Front. Genet.* **14**, 1252159 (2023).
41. Ahamed, B. S., Arya, M. S. & Nancy, V. A. O. Prediction of type-2 diabetes mellitus disease using machine learning classifiers and techniques. *Frontiers Comput. Science* **4** (2022).
42. Caballero-María, P. et al. Deep learning model approach to predict diabetes type 2 based on clinical, biochemical, and gut microbiota profiles. *Applied Sciences* **15** (2025).
43. Chen, T. et al. Association of waist-to-hip ratio adjusted for body mass index with cognitive impairment in middle-aged and elderly patients with type 2 diabetes mellitus: a cross-sectional study. *BMC Public. Health.* **24**, 2424 (2024).

44. Cho, S., Shin, A., Choi, J. Y., Lee, J. K. & Kang, D. Impact of body mass index and waist-to-hip ratio on mortality in middle-aged koreans: a prospective cohort study based on a health examinees study. *Epidemiol. Health.* **46**, e2024073 (2024).

45. Feng, X. et al. Correlation analysis of anthropometric indices and type 2 diabetes mellitus in residents aged 60 years and older. *Front. Public. Health.* **11**, 1122509 (2023).

46. Hasan, R., Dattana, V., Mahmood, S. & Hussain, S. Towards transparent diabetes prediction: Combining AutoML and explainable AI for improved clinical insights. *Information* **16** (2024).

47. Islam, M. M. et al. Explainable machine learning for efficient diabetes prediction using hyperparameter tuning, SHAP analysis, partial dependency, and LIME. *Engineering Reports* **7** (2024).

48. Kim, H. et al. National trends in type 2 diabetes mellitus stratified by central adiposity using waist-to-height ratio in South Korea, 2005–2022. *Sci. Rep.* **14**, 24273 (2024).

49. Liu, N. et al. Waist-to-hip ratio better reflect beta-cell function and predicts diabetes risk in adult with overweight or obesity. *Ann. Med.* **57**, 2462447 (2025).

50. Oumer, A. et al. Waist-to-hip circumference and waist-to-height ratio could strongly predict glycemic control than body mass index among adult patients with diabetes in ethiopia: ROC analysis. *PLOS ONE.* **17**, e0273786 (2022).

Tables

Table 1. Performance Comparison Across Machine Learning Models and Anthropometric Feature Configurations For Glucose Fasting Inclusion

Configuration	Model	CV AUC Mean	CV AUC SD	Validation AUC
BMI	LogReg	0.8862	0.0057	0.9115
BMI	LinearSVM	0.8866	0.0064	0.9090
BMI	RandomForest	0.8815	0.0080	0.8852
BMI	BalancedRF	0.8815	0.0108	0.8940
BMI	XGBoost	0.8854	0.0069	0.9015
BMI	CatBoost	0.8856	0.0117	0.9014
WHR	LogReg	0.8867	0.0060	0.9082
WHR	LinearSVM	0.8863	0.0068	0.9070
WHR	RandomForest	0.8814	0.0105	0.8820
WHR	BalancedRF	0.8802	0.0083	0.8908
WHR	XGBoost	0.8854	0.0099	0.8924
WHR	CatBoost	0.8839	0.0076	0.8910
BMI + WHR	LogReg	0.8884	0.0056	0.9122
BMI + WHR	LinearSVM	0.8884	0.0060	0.9100
BMI + WHR	RandomForest	0.8880	0.0112	0.8888
BMI + WHR	BalancedRF	0.8868	0.0100	0.8895
BMI + WHR	XGBoost	0.8883	0.0089	0.8963
BMI + WHR	CatBoost	0.8870	0.0106	0.8985
Abbreviations: BMI, body mass index; CV, cross-validation; ROC-AUC, area under the receiver operating characteristic curve; SD, standard deviation; SVM, support vector machine; WHR, waist-to-hip ratio; XGBoost, extreme gradient boosting.				

Table 2. Validation Performance of Deep Learning Models Across Anthropometric Feature Configurations

Configuration	Model	Validation ROC-AUC (Mean±SD)	Accuracy	Precision	Recall	Specificity	F1	Median Best Epoch
BMI	WideDeep	0.8897± 0.0038	0.8512	0.5541	0.7714	0.8681	0.6444	17
BMI	MLP	0.8891± 0.0027	0.8540	0.5602	0.7662	0.8725	0.6469	21
BMI	ResNet	0.8854± 0.0070	0.8231	0.4970	0.8000	0.8280	0.6127	6
WHR	WideDeep	0.8831± 0.0046	0.8367	0.5230	0.7558	0.8538	0.6178	17
WHR	MLP	0.8824± 0.0028	0.8413	0.5327	0.7688	0.8566	0.6287	16
WHR	ResNet	0.8794± 0.0065	0.8109	0.4757	0.7844	0.8165	0.5919	6
BMI+WHR	WideDeep	0.8868± 0.0048	0.8567	0.5658	0.7792	0.8731	0.6553	17
BMI+WHR	MLP	0.8862± 0.0052	0.8490	0.5482	0.7844	0.8626	0.6450	15
BMI+WHR	ResNet	0.8783± 0.0075	0.8145	0.4822	0.8104	0.8154	0.6043	5

Table 3: Final Test Performance of the Selected Deep-Learning Model

Performance Measure	Result
Model type	Deep learning
Configuration	BMI
Selected model	Wide & Deep
Probability threshold	0.20
Test ROC-AUC	0.9035
Accuracy	0.8294
Precision (PPV)	0.5035
Recall (Sensitivity)	0.7474
Specificity	0.8465
F1-score	0.6017
TN	386
FP	70
FN	24
TP	71
NPV	0.9415

Abbreviations: BMI, body mass index; FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; ROC-AUC, area under the receiver operating characteristic curve; TN, true negative; TP, true positive.

Table 4. Global SHAP Feature Importance for the Final Wide & Deep Model Using the BMI Configuration in the Male-Only Cohort

Rank	Feature	Mean SHAP
1	Age	0.0809
2	Fasting glucose Level	0.0734
3	Family history diabetes	0.0612
4	Hypertension history	0.0419
5	Diastolic blood pressure	0.0390
6	Heart rate	0.0388
7	LDL cholesterol	0.0356
8	Non-Hispanic White ethnicity	0.0301
9	Body mass index (BMI)	0.0295
10	HDL cholesterol	0.0208
11	College graduate or above	0.0166
12	Non-Hispanic Black ethnicity	0.0164
13	Sleep duration (hours/day)	0.0160
14	Some college/Associate degree	0.0147
15	Poverty-income ratio	0.0127
16	High school graduate/GED	0.0125
17	Physical activity (minutes/week)	0.0113
18	Systolic blood pressure	0.0100
19	Non-Hispanic Asian ethnicity	0.0097
20	Other Hispanic ethnicity	0.0092
21	Ever smoked 100 cigarettes	0.0090
22	Total cholesterol	0.0087
23	Triglycerides	0.0074
24	Alcohol drinks per drinking day	0.0061
25	Less than 9th grade education	0.0022

Abbreviations: BMI, body mass index; GED, General Educational Development; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SHAP, SHapley Additive exPlanations.

Figures

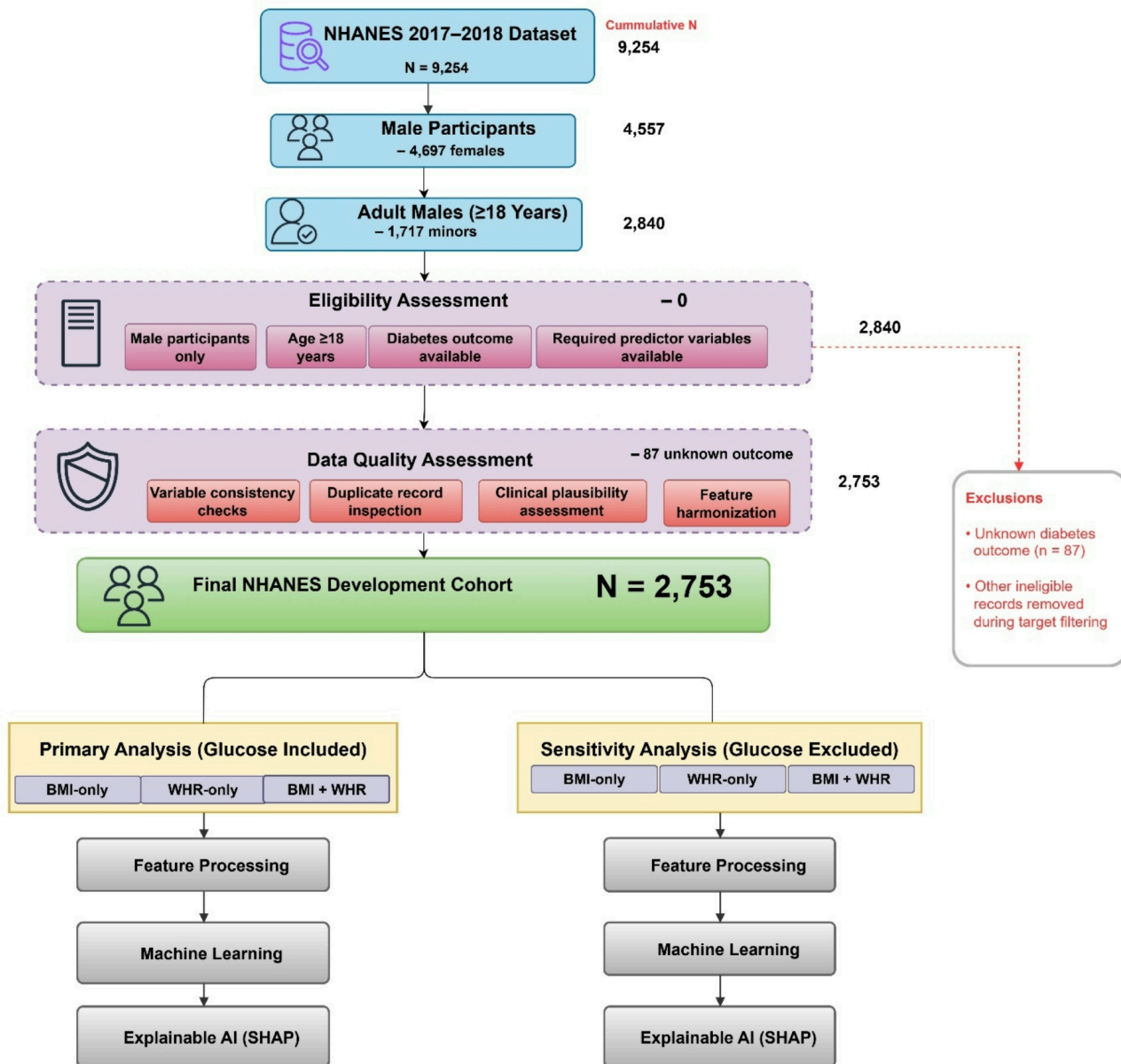


Figure 1

Study Cohort Selection and Analytical Workflow for the NHANES 2017–2018 Male Cohort. The NHANES 2017–2018 dataset initially included 9,254 participants and after restriction to male participants and adults aged ≥18 years, 2,840 participants remained. Following outcome and data-quality assessment, 87 participants with unknown diabetes status were excluded, resulting in a final development cohort of 2,753 adult males. The analytical framework comprised a primary analysis including fasting glucose and a sensitivity analysis excluding fasting glucose. For both, the BMI-only, WHR-only and BMI+WHR feature set combinations were analyzed. Afterwards, the feature processing, the machine-learning modeling, and the SHAP-based model explainability analysis were performed.

Abbreviations: BMI, body mass index; NHANES, National Health and Nutrition Examination Survey; SHAP, SHapley Additive exPlanations; WHR, waist-to-hip ratio.

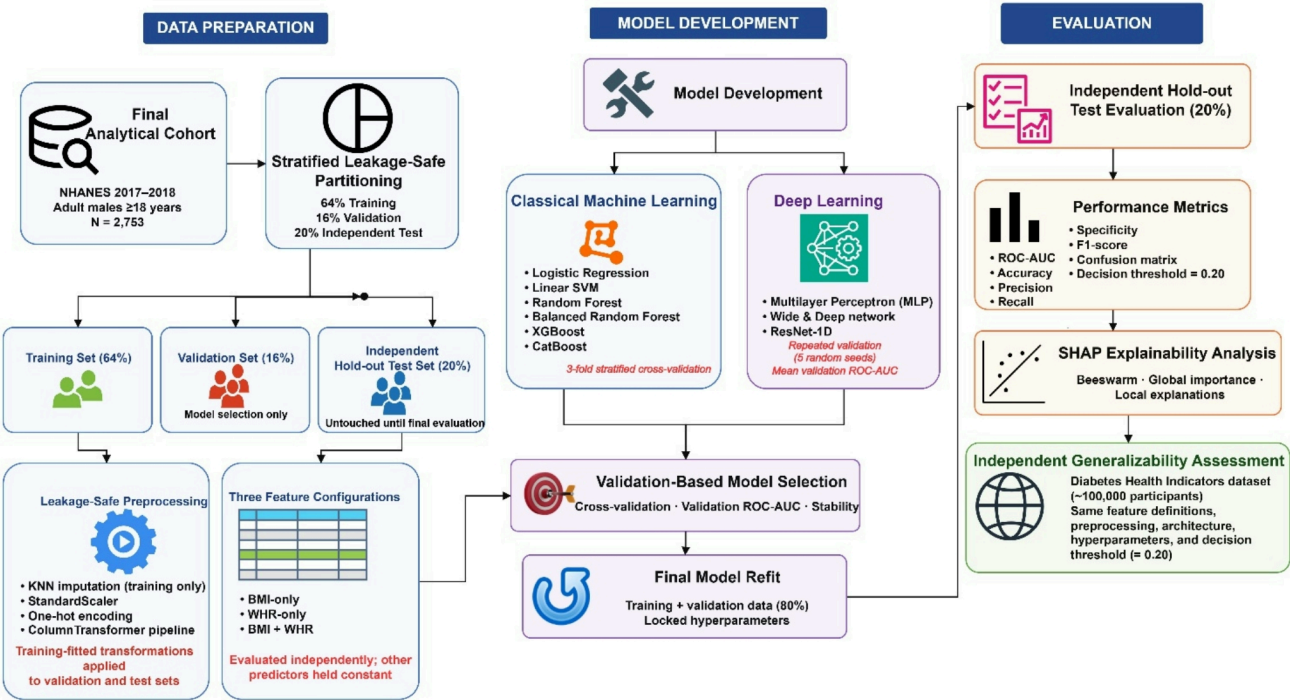


Figure 2

Leakage-Safe Machine Learning Framework for Model Development, Evaluation, and Explainability. The final analytical cohort was partitioned using stratified leakage-safe splitting into training (64%), validation (16%), and independent hold-out test (20%) sets. Preprocessing transformations were fitted using the training data and subsequently applied to the validation and test sets. BMI-only, WHR-only, and combined BMI+WHR anthropometric feature configurations were evaluated using classical machine-learning and deep-learning models. Model selection was performed using validation performance, after which the selected model was refitted using the combined training and validation data with locked hyperparameters. Final performance was evaluated on the untouched independent hold-out test set using a prespecified decision threshold of 0.20. SHAP was used to evaluate model's interpretability, and the generalizability of the model was further independently evaluated in Diabetes Health Indicators.

Abbreviations: BMI, body mass index; KNN, k-nearest neighbors; MLP, multilayer perceptron; ResNet-1D, one-dimensional residual network; ROC-AUC, area under the receiver operating characteristic curve; SHAP, SHapley Additive exPlanations; SVM, support vector machine; WHR, waist-to-hip ratio.

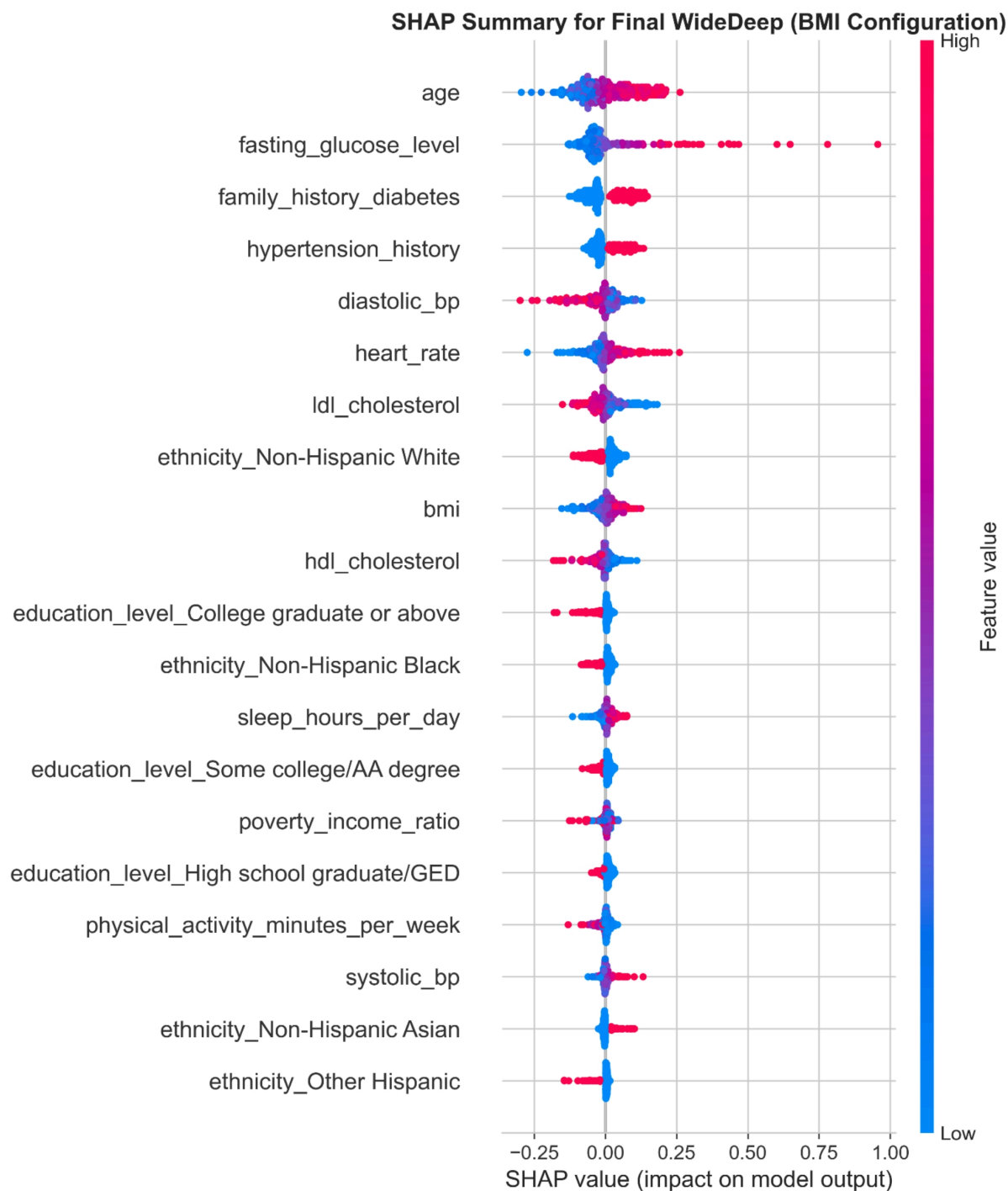


Figure 3

SHAP Summary Plot for the Final Wide & Deep Model Using the BMI Configuration. The SHAP beeswarm plot illustrates the global contribution of individual features to the predictions of the final Wide & Deep model. Features are sorted in descending order of total contribution as measured by the average absolute SHAP value. Points are individual participants, horizontally located at their SHAP value, colored according to their feature value, low (blue) to high (red). A positive SHAP value contributes positively to higher model output, while a negative SHAP value contributes negatively to lower model output.

Abbreviations: AA, associate degree; BMI, body mass index; BP, blood pressure; GED, General Educational Development; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SHAP, SHapley Additive exPlanations.



Figure 4

SHAP Summary Plot for the Final Linear Support Vector Machine Model in the Glucose-Excluded Analysis. The SHAP beeswarm plot illustrates the global contributions of predictor variables to the output of the selected linear support vector machine model using the combined BMI and WHR configuration after exclusion of fasting glucose.

The left-to-right order of the points in each plot is determined by their average absolute SHAP value (most to least important feature). Points are colored from (blue) low to (red) high by their feature value. The x-axis is their SHAP value which for these plots is the contribution of each feature to the predicted value. Positive SHAP values represent contributions of the feature to a higher predicted value. Negative SHAP values represent contributions to a lower predicted value. The bar plot to the right shows the absolute SHAP values of the features averaged for all participants. The feature colors correspond to their color in the scatterplot.

Abbreviations: BMI, body mass index; BP, blood pressure; GED, General Educational Development; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SHAP, SHapley Additive exPlanations; SVM, support vector machine; WHR, waist-to-hip ratio.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [OnlineOnlySupplementalMaterial.pdf](#)