SUPLLEMENTARY TABLES

Supplementary Table S1. Summary of Prior Studies Supporting Anthropometric Comparison and Explainable Machine Learning

Study	Data / population (as reported)	Models/ Methods used	Explainability Approach	Key takeaway aligned to this study
Seo BM et al., 2024	Medical examination data (population-based)	ML models including boosting (e.g., XGBoost)	Feature importance / interpretable predictors emphasized	Shows boosting-style models can identify key diabetes predictors and handle mixed clinical variables well.
Li et al., 2024	Diabetes prediction with class imbalance handling	GA-XGBoost + SMOTE	SHAP used to address black-box concerns	Directly motivates pairing strong ensembles with SHAP because performance alone is insufficient clinically.
Lee et al., 2025	Older adults in Korea (risk prediction)	ML models including XGBoost	SHAP explicitly described for feature contribution	Supports SHAP as a standard, defensible interpretability method in diabetes risk models.
Tasin et al., 2023	Review-style synthesis of ML and XAI for diabetes	Multiple ML methods across studies	Notes lack of explainability in prior works	Reinforces the research need: many studies optimize accuracy but under-deliver on explainability.
Srinivasu et al., 2024	AI system for diabetes prediction (healthcare context)	Deep learning (black-box) + XAI	SHAP described as feature-contribution explanation	Strong justification that SHAP is needed to make black-box models clinically transparent.
Sadeghi et al., 2024	Male vs female ROC comparisons of anthropometric indices	Comparative ROC for BMI vs central indices	Not ML-focused (clinical discrimination)	Reports WHR/WHtR/BRI-type indices discriminate diabetes better than BMI in men → supports your BMI vs WHR comparison rationale.
Kobayashi et al., 2024	Sex-stratified diabetes incidence models	Anthropometric indices compared	Not ML-focused	Explicitly highlights the unresolved “best index” problem and details BMI limitations (no fat distribution).

Abbreviations: BMI, body mass index; BRI, body roundness index; GA, genetic algorithm; ML, machine learning; ROC, receiver operating characteristic; SHAP, SHapley Additive exPlanations; SMOTE, synthetic minority oversampling technique; WHR, waist-to-hip ratio; WHtR, waist-to-height ratio; XAI, explainable artificial intelligence; XGBoost, extreme gradient boosting.

Supplementary Table S2. Prior Machine-Learning Studies Motivating Explainability and Anthropometric Comparison in Diabetes Risk Prediction

Study	Core ML approach	Imbalance handling	Explainability	What it supports in this paper
Rafie et al., 2025	ML based diabetes risk prediction	Discusses imbalance handling needs	Explicitly prioritizes SHAP/XAI	ML is needed because traditional tools miss complex interactions; interpretability is required for clinical usability
Li et al., 2024	GA-optimized XGBoost	SMOTE	SHAP	Prior work pairs tree ensembles with oversampling + SHAP; motivates model-agnostic explainability
Tasin et al., 2023	XGBoost and ADASYN	ADASYN	SHAP and LIME	Diabetes prediction studies use resampling + XAI; highlights importance of explainability
Ejiyi OS et al., 2025	CatBoost	SMOTE	SHAP	Clinical barrier is black-box behavior; SHAP is positioned as remedy
Sun et al., 2024	Comparative ROC analysis (men vs women)	Not applicable	Not applicable	Provides evidence for male-specific differences in the discriminatory performance of BMI and central adiposity measures (WHR/WHtR)

Abbreviations: ADASYN, adaptive synthetic sampling; BMI, body mass index; GA, genetic algorithm; LIME, local interpretable model-agnostic explanations; ML, machine learning; ROC, receiver operating characteristic; SHAP, SHapley Additive exPlanations; SMOTE, synthetic minority oversampling technique.

Supplementary Table S3. Anthropometric Feature Configurations Evaluated

Configuration	What's included	What's excluded	Purpose
BMI	Full predictor set + BMI	WHR	BMI-only anthropometric signal
WHR	Full predictor set + WHR	BMI	WHR-only anthropometric signal
BOTH	Full predictor set, BMI, WHR	None	Test incremental value / complementarity

Supplementary Table S4. Baseline Characteristics and Feature Domains Included in the Male-Only Diabetes Prediction Dataset

Category	Variables (raw feature names)	Handling in pipeline
Demographic	age	Numeric (standardized)
Anthropometric (core comparison)	bmi, waist_to_hip_ratio	Numeric (standardized); used in BMI-only, WHR-only, and BMI+WHR feature configurations
Blood pressure & cardiovascular	systolic_bp, diastolic_bp, heart_rate, hypertension_history	Numerical variables standardized; hypertension_history one-hot encoded
Lipid profile	cholesterol_total, hdl_cholesterol, ldl_cholesterol, triglycerides	Numeric (standardized)
Glycemic indicator	glucose_fasting	Numeric (standardized); retained as a predictor and used during outcome quality-control procedures
Lifestyle & behavioral factors	physical_activity_minutes_per_week, alcohol_drinks_per_drinking_day, sleep_hours_per_day, ever_smoked_100_cigarettes	Numerical variables standardized; smoking status one-hot encoded
Socioeconomic factors	education_level, poverty_income_ratio	poverty_income_ratio standardized; education_level one-hot encoded
Clinical history	family_history_diabetes	One-hot encoded
Demographic background	ethnicity	One-hot encoded

Abbreviations: BMI, body mass index; BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein; WHR, waist-to-hip ratio.

Supplementary Table S5. Threshold-Based Performance Metrics (Probability Threshold = 0.20) For the Best Classical Model for Glucose Inclusion in Each Anthropometric Feature Configuration

Configuration	Best Model	Precision (PPV)	Recall (Sensitivity)	Specificity	F1-score	TN	FP	FN	TP	NPV
BMI	LinearSVM	0.5413	0.7662	0.8626	0.6344	314	50	18	59	0.9458
WHR	LogReg	0.3380	0.9481	0.6071	0.4983	221	143	4	73	0.9822
BMI + WHR	LinearSVM	0.5446	0.7922	0.8599	0.6455	313	51	16	61	0.9514

Abbreviations: BMI, body mass index; FN, false negative; FP, false positive; LinearSVM, linear support vector machine; LogReg, logistic regression; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive; WHR, waist-to-hip ratio.

Supplementary Table S6. Classical Machine Learning Model Selection Without Fasting Glucose

Configuration	Model	CV AUC Mean± SD	Validation AUC
BMI	Logistic Regression	0.8619 ± 0.0090	0.8784
BMI	LinearSVM	**0.8621 ± 0.0098**	0.8772
BMI	Random Forest	0.8519 ± 0.0021	0.8634
BMI	Balanced Random Forest	0.8592 ± 0.0032	0.8677
BMI	XGBoost	0.8542 ± 0.0059	0.8617
BMI	CatBoost	0.8479 ± 0.0029	0.8718
WHR	Logistic Regression	0.8637 ± 0.0103	0.8773
WHR	LinearSVM	**0.8643 ± 0.0112**	0.8765
WHR	Random Forest	0.8557 ± 0.0045	0.8509
WHR	Balanced Random Forest	0.8583 ± 0.0057	0.8565
WHR	XGBoost	0.8588 ± 0.0053	0.8568
WHR	CatBoost	0.8540 ± 0.0016	0.8662
BMI + WHR	Logistic Regression	0.8654 ± 0.0089	**0.8808**
BMI + WHR	**LinearSVM**	**0.8658 ± 0.0103**	**0.8807**
BMI + WHR	Random Forest	0.8568 ± 0.0047	0.8518
BMI + WHR	Balanced Random Forest	0.8604 ± 0.0079	0.8695
BMI + WHR	XGBoost	0.8592 ± 0.0052	0.8605
BMI + WHR	CatBoost	0.8573 ± 0.0012	0.8615

Abbreviations: BMI, body mass index; CV, cross-validation; ROC-AUC, area under the receiver operating characteristic curve; SVM, support vector machine; WHR, waist-to-hip ratio; XGBoost, extreme gradient boosting.

Supplementary Table S7. Deep Learning Model Selection Without Fasting Glucose

Configuration	Model	Validation ROC-AUC (mean ± SD)[a]	Accuracy	Precision	Recall	Specificity	F1	Median Best Epoch
BMI	**MLP**	**0.8573 ± 0.0029**	0.8154	0.4807	0.7039	0.8390	0.5712	17
BMI	WideDeep	0.8562 ± 0.0044	0.8163	0.4821	0.6857	0.8440	0.5658	17
BMI	ResNet	0.8465 ± 0.0068	0.7986	0.4535	0.7195	0.8154	0.5554	5
WHR	WideDeep	**0.8464 ± 0.0019**	0.8095	0.4700	0.7117	0.8302	0.5659	16
WHR	MLP	0.8459 ± 0.0025	0.8041	0.4620	0.7247	0.8209	0.5639	17
WHR	ResNet	0.8407 ± 0.0047	0.7791	0.4264	0.7506	0.7852	0.5431	5
BMI + WHR	**MLP**	**0.8547 ± 0.0024**	0.8177	0.4855	0.7013	0.8423	0.5734	18
BMI + WHR	WideDeep	0.8538 ± 0.0027	0.8154	0.4834	0.7169	0.8363	0.5760	12
BMI + WHR	ResNet	0.8503 ± 0.0028	0.7968	0.4513	0.7429	0.8082	0.5612	5

[a]Bold values indicate the highest mean validation ROC-AUC within each anthropometric feature configuration.

Supplementary Table S8. Threshold-Based Performance Metrics (Probability Threshold = 0.20) Performance Comparison of the Best Glucose-Excluded Models Across BMI, WHR, and BMI + WHR Feature Configurations.

Configuration	Best Model	Precision (PPV)	Recall	Specificity	F1-score	TN	FP	FN	TP	NPV
BMI	LinearSVM	0.4590	0.7273	0.8187	0.5628	298	66	21	56	09341
WHR	LinearSVM	0.4836	0.7662	0.8269	0.5930	301	63	18	59	0.9435
BMI + WHR	LinearSVM	0.4836	0.7662	0.8269	0.5930	301	63	18	59	0.9435

FOR INDEPENDENT VALIDATION

Supplementary Table S9. Baseline Characteristics and Feature Domains Included in the Male-Only Diabetes Prediction Dataset

Category	**Variables (raw Feature names)**	**Handling in pipeline**
Demographic	age	Numeric (standardized)
Anthropometric (core comparison)	bmi, waist_to_hip_ratio	Numeric (standardized)- used in BMI-only / WHR-only / BOTH configurations
Blood pressure / vitals	systolic_bp, diastolic_bp, heart_rate, hypertension_history	Numeric (standardized)
Lipid profile	cholesterol_total, hdl_cholesterol, ldl_cholesterol, triglycerides	Numeric (standardized)
Glycemic indicator (used for cleaning and prediction)	glucose_fasting	Numeric (standardized)-retained as predictor and used in label-cleaning step
Lifestyle / behavior	physical_activity_minutes_per_week, alcohol_consumption_per_week, diet_score, sleep_hours_per_day, screen_time_hours_per_day	Numeric (standardized)

Supplementary Table S10. Performance Comparison Across Machine Learning Models and Anthropometric Feature Configurations

Configuration	Model	CV AUC mean	CV AUC SD	Validation AUC
BMI	LogReg	0.8543	0.0077	0.8733
BMI	LinearSVM	0.8543	0.0078	0.8731
BMI	RandomForest	0.8447	0.0084	0.8612
BMI	BalancedRF	0.8437	0.0075	0.8617
BMI	XGBoost	0.8525	0.0077	0.8707
BMI	CatBoost	0.8499	0.0073	0.8722
WHR	LogReg	0.8545	0.0075	0.8730
WHR	LinearSVM	0.8546	0.0074	0.8727
WHR	RandomForest	0.8436	0.0070	0.8605
WHR	BalancedRF	0.8435	0.0068	0.8647
WHR	XGBoost	0.8521	0.0072	0.8705
WHR	CatBoost	0.8501	0.0056	0.8742
BMI + WHR	LogReg	0.8544	0.0075	0.8730
BMI + WHR	LinearSVM	0.8543	0.0075	0.8728
BMI + WHR	RandomForest	0.8446	0.0075	0.8604
BMI + WHR	BalancedRF	0.8426	0.0082	0.8630
BMI + WHR	XGBoost	0.8523	0.0073	0.8710
BMI + WHR	CatBoost	0.8497	0.0066	0.8716

Supplementary Table S11. Validation Performance of Deep Learning Models Across Anthropometric Feature Configurations

Configuration	Model	Validation ROC-AUC (Mean ± SD)	Accuracy	Precision	Recall	Specificity	F1-score	Median Best Epoch
BMI	**MLP**	**0.8726 ± 0.0012**	0.5907	0.5609	0.9810	0.1681	0.7137	5
BMI	ResNet	0.8707 ± 0.0013	0.6269	0.5866	0.9602	0.2660	0.7281	3
BMI	WideDeep	0.8702 ± 0.0017	0.6229	0.5832	0.9637	0.2539	0.7266	7
WHR	**MLP**	**0.8722 ± 0.0017**	0.5936	0.5627	0.9802	0.1749	0.7150	4
WHR	ResNet	0.8697 ± 0.0018	0.6344	0.5918	0.9584	0.2835	0.7316	4
WHR	WideDeep	0.8694 ± 0.0018	0.6232	0.5834	0.9633	0.2549	0.7266	5
BMI + WHR	**MLP**	**0.8715 ± 0.0011**	0.5927	0.5624	0.9808	0.1726	0.7148	5
BMI + WHR	WideDeep	0.8694 ± 0.0028	0.6223	0.5833	0.9604	0.2563	0.7256	6
BMI + WHR	ResNet	0.8685 ± 0.0019	0.6275	0.5871	0.9567	0.2710	0.7276	3

Supplementary Table S12. Threshold-Based Metrics at Probability Threshold 0.20 for the Best Classical Model in Each Configuration with Fasting Glucose

Configuration	Best Model	Precision (PPV)	Recall (Sensitivity)	Specificity	F1-score	TN	FP	FN	TP	NPV
BMI	LogReg	0.6123	0.9388	0.3565	0.7412	603	1,088	112	1,719	0.8433
WHR	LogReg	0.6112	0.9377	0.3542	0.7400	599	1,092	114	1,717	0.8401
BMI + WHR	LogReg	0.6107	0.9366	0.3536	0.7393	598	1,093	116	1,715	0.8375

Supplementary Table S13. Global SHAP Feature Importance for Final Deep-Learning Model in the Male-Only Independent Cohort)

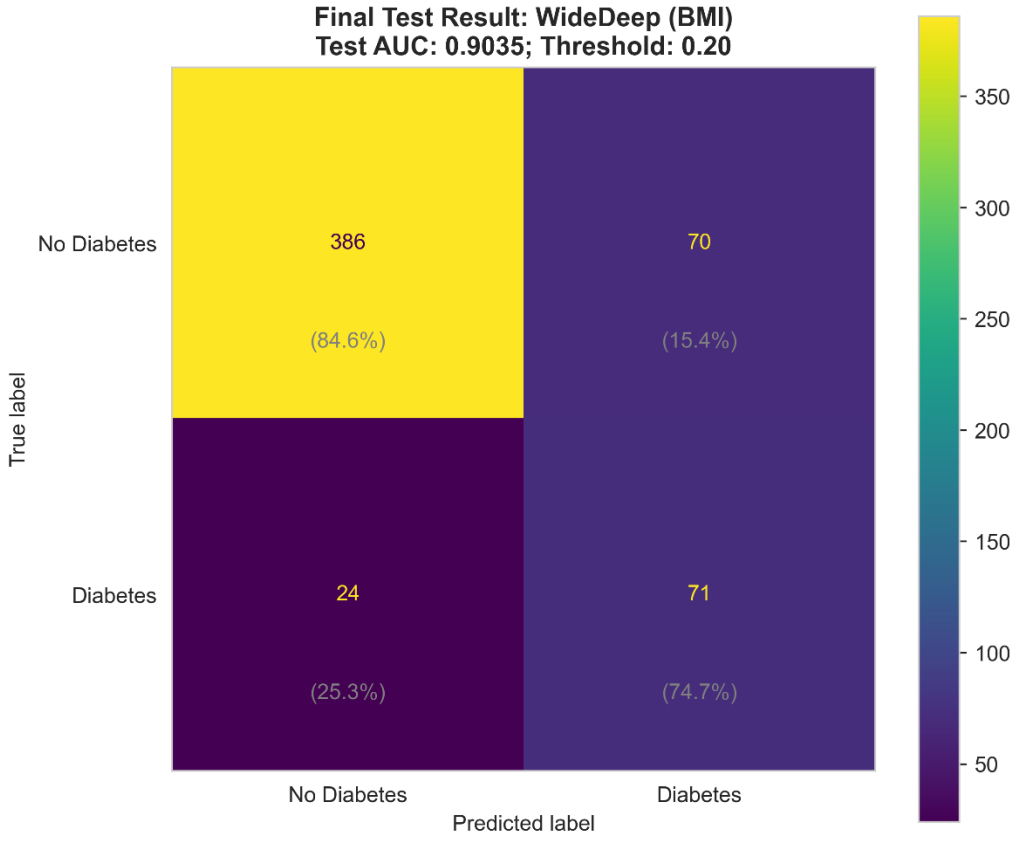
Rank	Feature	Mean |SHAP|
1	glucose_fasting	0.257186
2	family_history_diabetes_1	0.021598
3	age	0.018105
4	physical_activity_minutes_per_week	0.016096
5	screen_time_hours_per_day	0.007355
6	bmi	0.007289
7	cholesterol_total	0.007252
8	hdl_cholesterol	0.006463
9	ldl_cholesterol	0.006187
10	diet_score	0.006154
11	heart_rate	0.005790
12	alcohol_consumption_per_week	0.005392
13	diastolic_bp	0.005332
14	income_level_Middle	0.004697
15	ethnicity_White	0.004605
16	sleep_hours_per_day	0.004365
17	ethnicity_Black	0.004089
18	systolic_bp	0.003972
19	income_level_Low	0.003698
20	ethnicity_Hispanic	0.003247
21	triglycerides	0.003148
22	hypertension_history_1	0.002836
23	income_level_Lower-Middle	0.002609
24	education_level_No formal	0.002310
25	education_level_Highschool	0.002244

Supplementary Table S14. Threshold Based Metrics Performance Comparison Of the Best Glucose-Excluded Models Across BMI, WHR, and BMI + WHR Feature Configurations.

Configuration	Model	CV ROC-AUC, mean ± SD	Validation ROC-AUC	Accuracy	Precision	Recall (Sensitivity)	Specificity	F1-score	TN	FP	FN	TP	NPV
BMI	Linear SVM	0.7475 ± 0.0073	**0.7411**	0.5454	0.5342	0.9820	0.0727	0.6919	123	1,568	33	1,798	0.7885
BMI + WHR	Linear SVM	0.7475 ± 0.0075	0.7409	0.5454	0.5341	**0.9831**	0.0716	**0.6922**	121	1,570	**31**	**1,800**	**0.7961**
WHR	Linear SVM	0.7463 ± 0.0078	0.7405	0.5429	0.5328	0.9814	0.0680	0.6906	115	1,576	34	1,797	0.7718

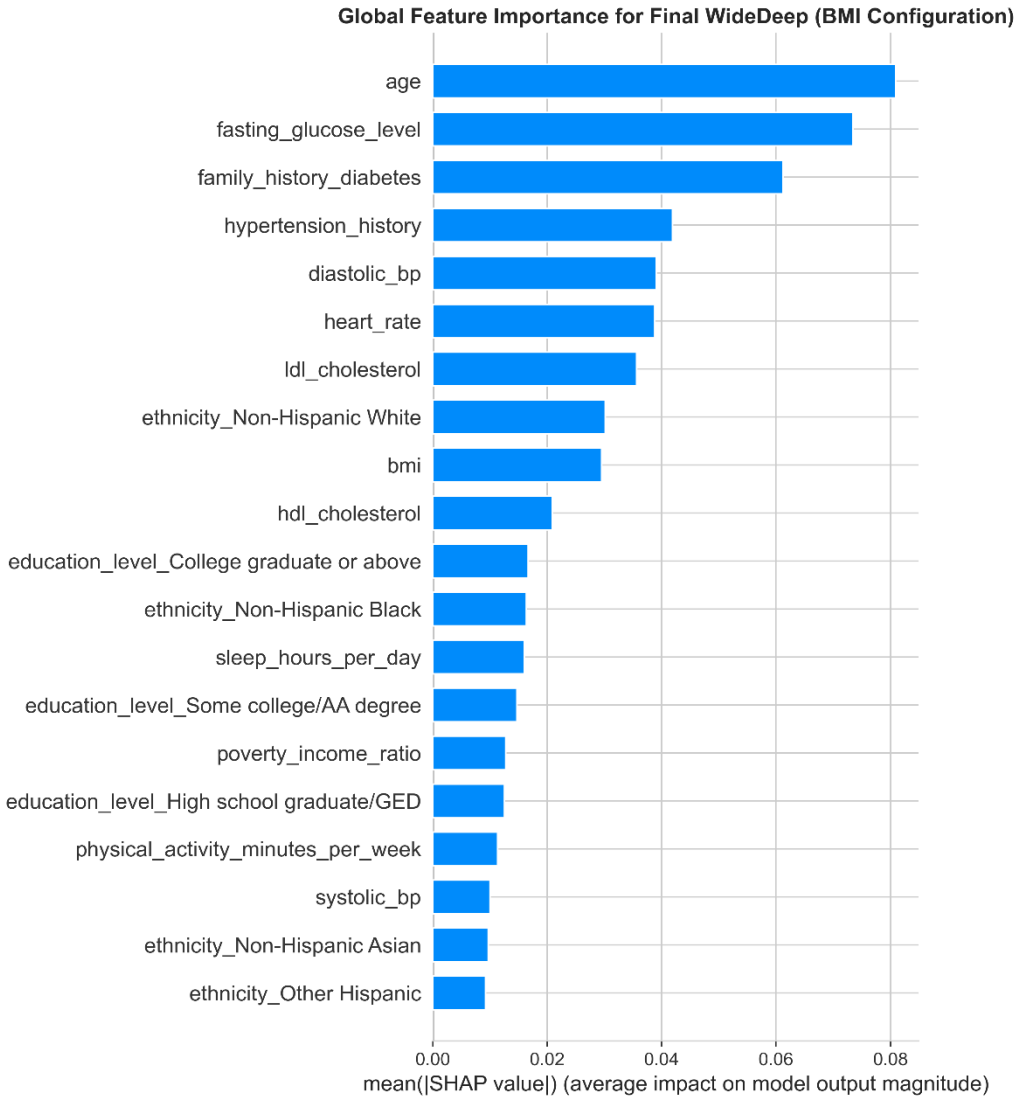
Supplementary Figure S1. Confusion Matrix of the Final Wide & Deep Model Using the BMI Configuration. Performance of the optimal Wide & Deep model in the independent hold-out test set measured by a confusion matrix for a probability threshold of 0.20. This Optimal model accurately predicted 386 non-diabetic and 71 diabetic participant, 70 participants without diabetes were classified as having diabetes and 24 diabetic participants were classified as not having diabetes. Test ROC-AUC is 0.9035.

Abbreviations: AUC, area under the receiver operating characteristic curve; BMI, body mass index; ROC-AUC, area under the receiver operating characteristic curve.



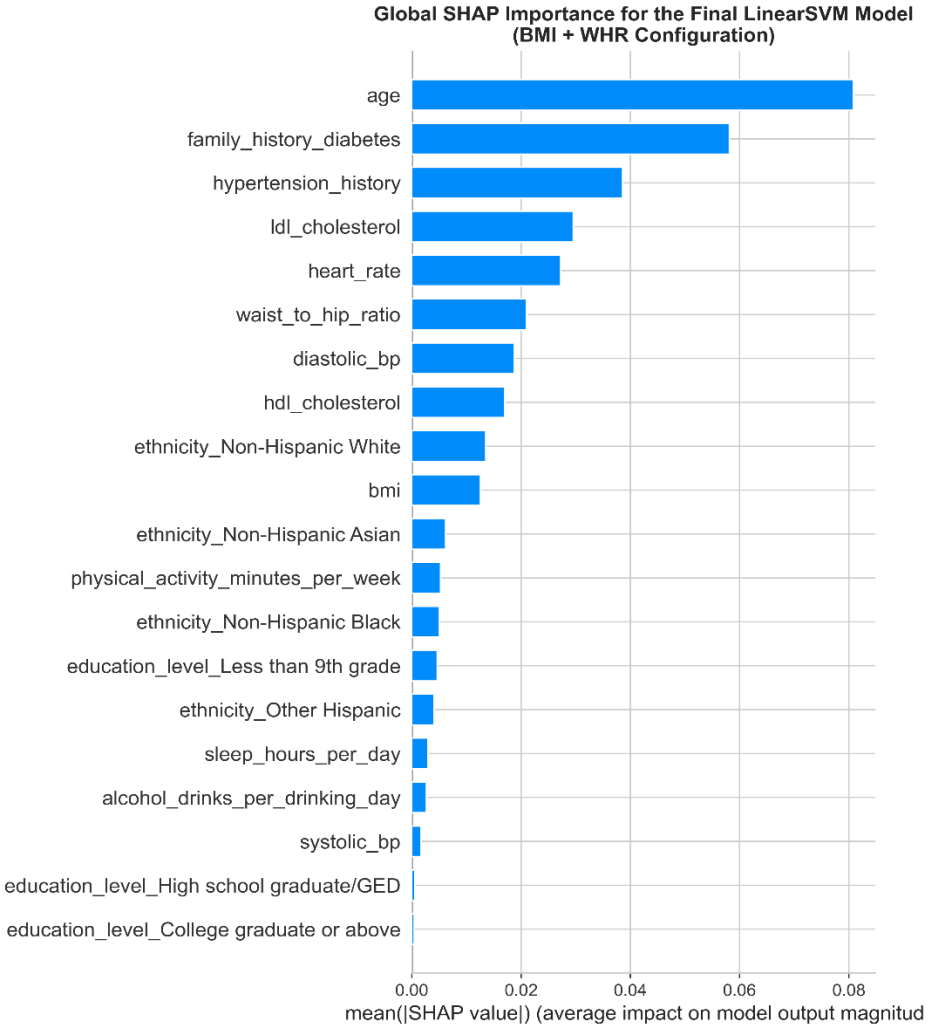
Supplementary Figure S2. Global SHAP Feature Importance for the Final Wide & Deep Model Using the BMI Configuration. The bar plot shows the global importance of the leading predictor variables based on their mean absolute SHAP values. Features are ranked in descending order of importance, with larger mean absolute SHAP values indicating a greater average contribution to the model output across participants. Age, fasting glucose level, family history of diabetes, and hypertension history were among the most influential predictors.

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.

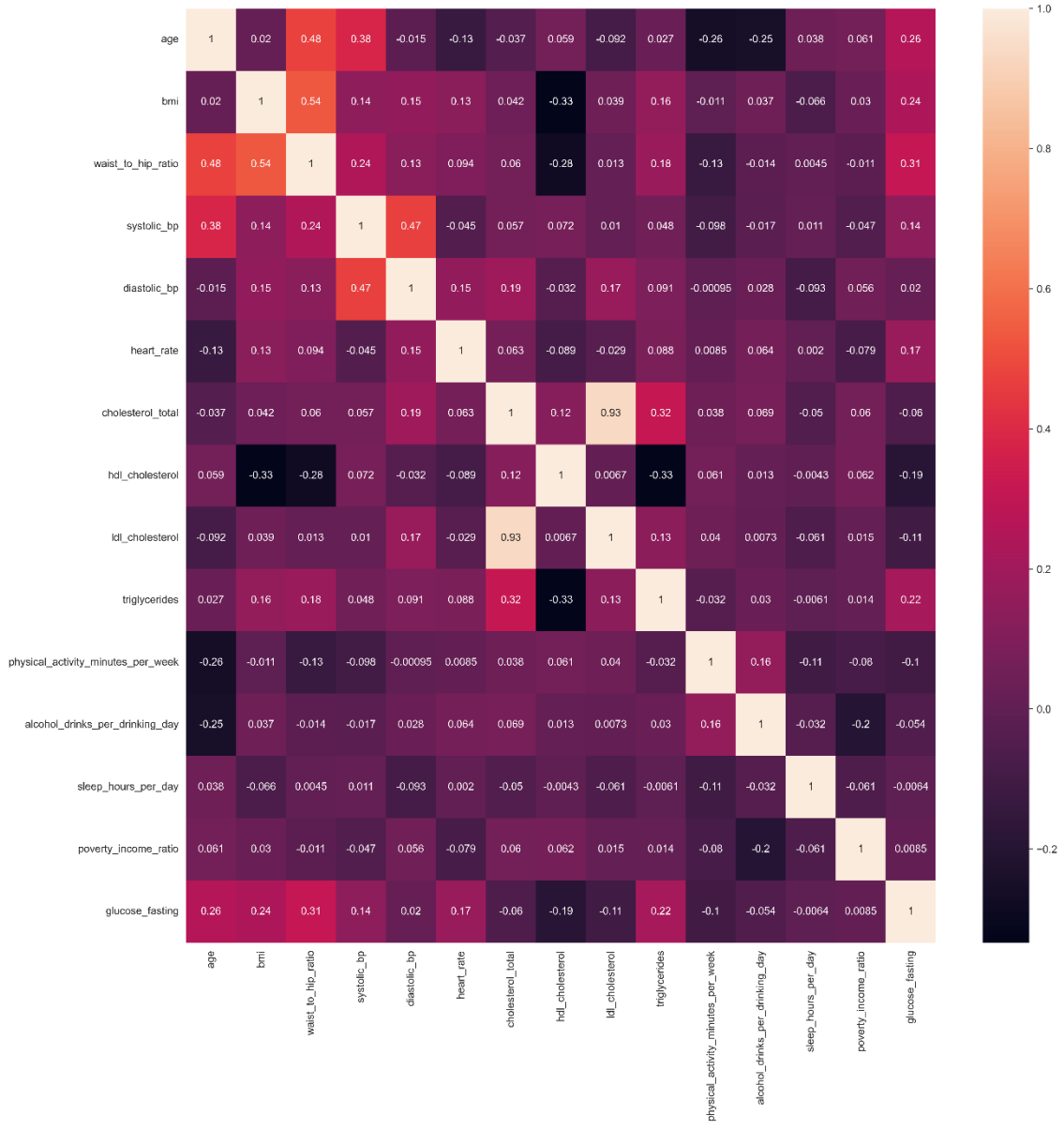


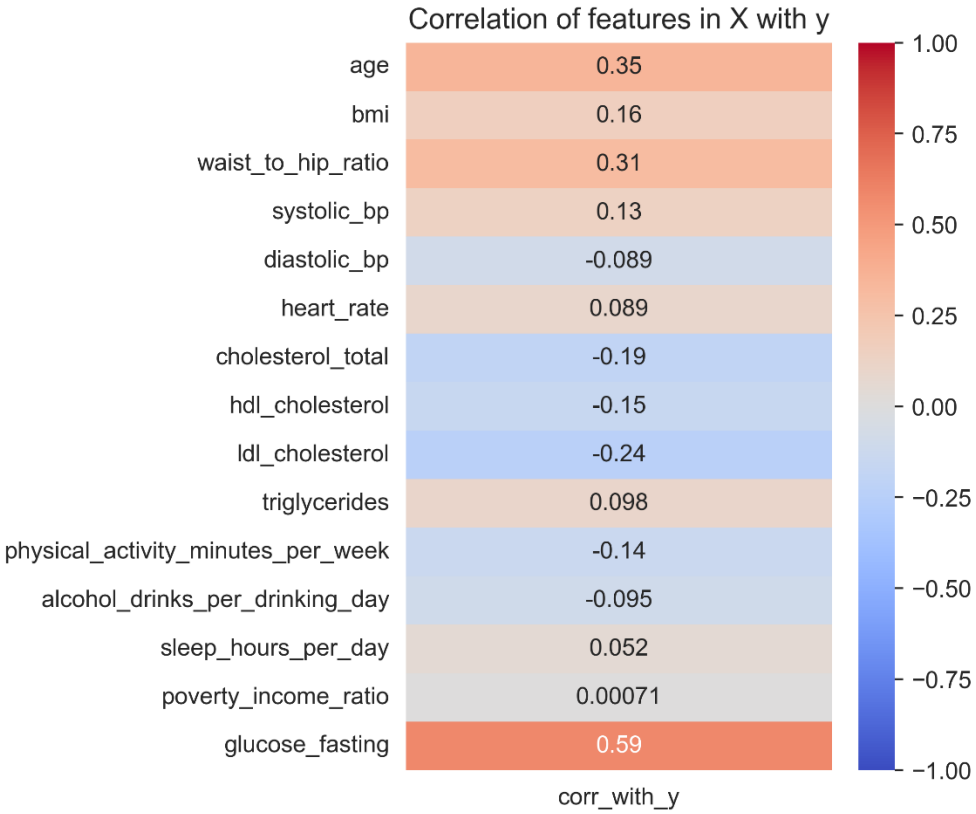
Supplementary Figure S3. Global SHAP Feature Importance for the Final Linear Support Vector Machine Model in the Glucose-Excluded Analysis. The bar plot presents global feature importance based on the mean absolute SHAP values for the selected linear support vector machine model using the combined BMI and WHR configuration after exclusion of fasting glucose. Features are ranked in descending order according to their mean absolute SHAP values, representing their average contribution to the magnitude of the model output. Larger mean absolute SHAP values indicate greater overall influence on model predictions.

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.

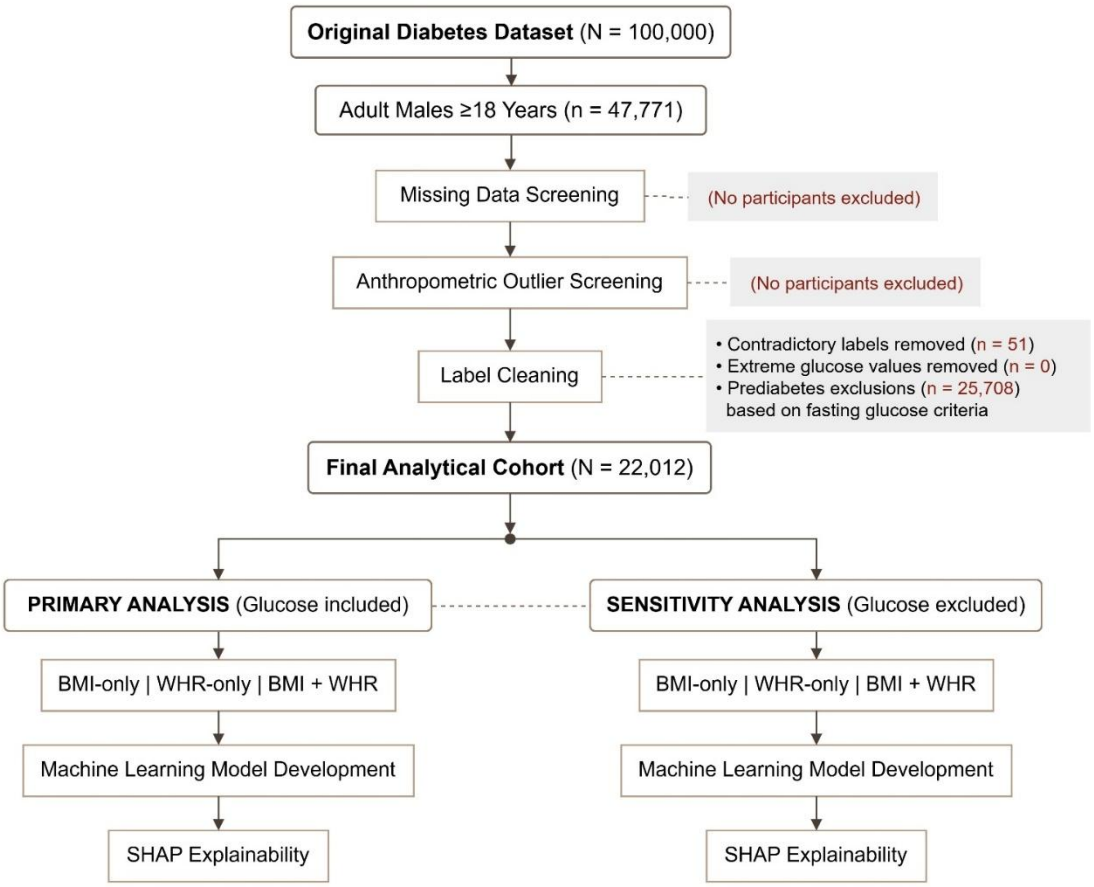


Supplementary Figure S4. Correlation Plots of Predictor Variables in the Male-Only Analytical Cohort. (A) Plots of the two-way Pearson correlation coefficient between continuous predictor variables. (B) Plot of Pearson correlation coefficients of individual predictor variables with binary diabetes status. Positive coefficients indicate positive associations, whereas negative coefficients indicate inverse associations. The magnitude of the coefficient represents the strength of the linear association. **Abbreviations:** BMI, body mass index; BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein; WHR, waist-to-hip ratio.

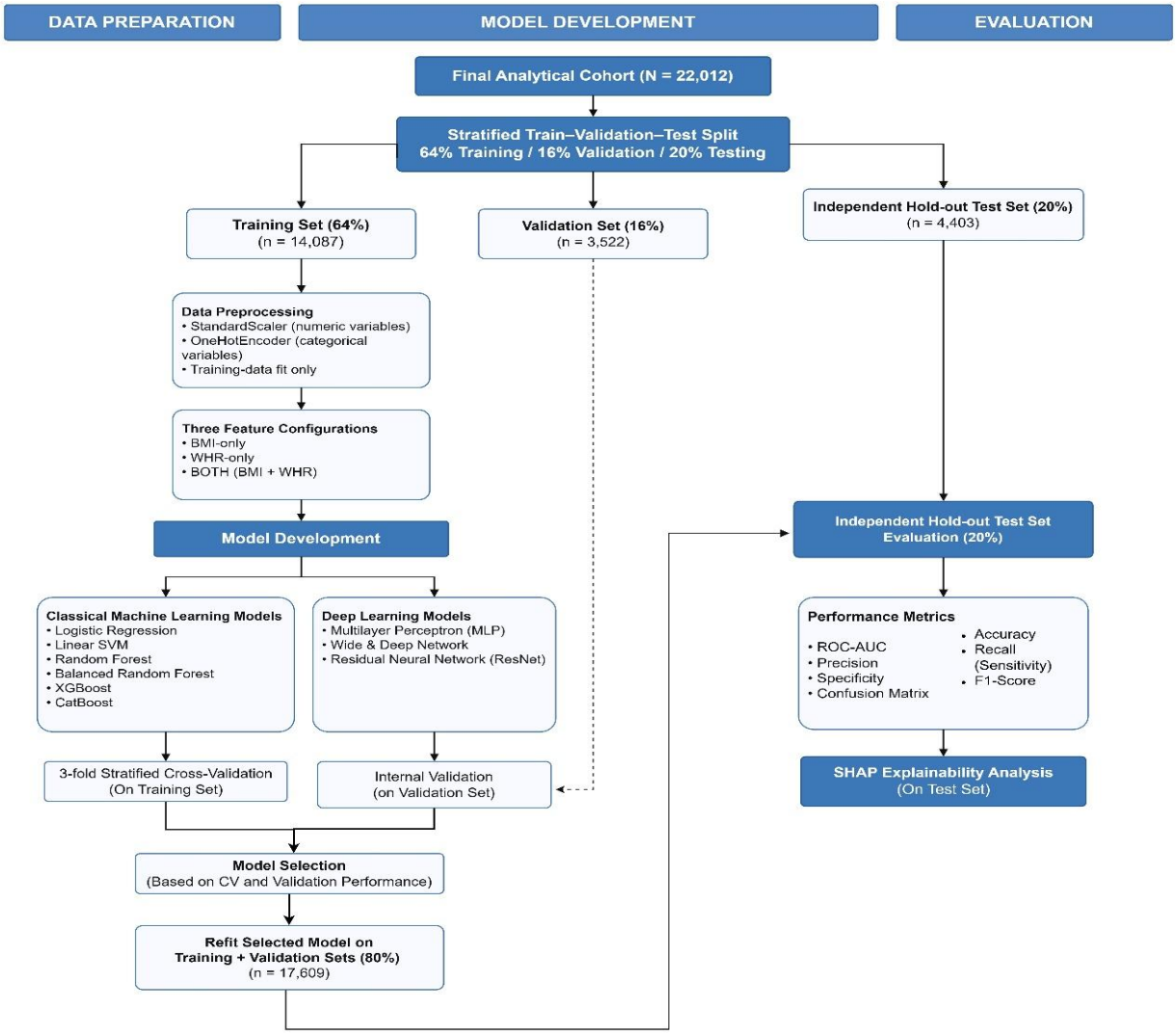



Correlation of features in X with y
age
bmi
waist_to_hip_ratio
systolic_bp
diastolic_bp
heart_rate
cholesterol_total
hdl_cholesterol
ldl_cholesterol
triglycerides
physical_activity_minutes_per_week
alcohol_drinks_per_drinking_day
sleep_hours_per_day
poverty_income_ratio
glucose_fasting
0.35
0.16
0.31
0.13
-0.089
0.089
-0.19
-0.15
-0.24
0.098
-0.14
-0.095
0.052
0.00071
0.59
corr_with_y
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Supplementary Figure S5. Population Selection Flowchart for the Independent Validation Cohort. The Diabetes Health Indicators dataset initially included 100,000 participants and after restriction to adult males aged ≥18 years, 47,771 participants remained. This is followed by label cleaning, including removal of contradictory diabetes labels and exclusion of participants meeting the predefined prediabetes criteria, the final analytical cohort comprised 22,012 participants. The cohort was subsequently evaluated under glucose-inclusive primary analysis and glucose-excluded sensitivity analysis using BMI-only, WHR-only, and combined BMI+WHR feature configurations, followed by machine-learning model development and SHAP-based explainability analysis. **Abbreviations:** BMI, body mass index; SHAP, SHapley Additive exPlanations; WHR, waist-to-hip ratio.



Supplementary Figure S6. Methodological Workflow for Independent Validation. The final analytical cohort (N=22,012) was stratified into an 80% training set and a 20% independent hold-out test set. Data preprocessing was fitted using the training data and included standardization of numerical variables and one-hot encoding of categorical variables. Three anthropometric feature configurations BMI-only, WHR-only, and combined BMI+WHR were evaluated. Classical machine-learning algorithms ised in this analysis were logistic regression, linear support vectormachine, randomforest, balanced random forest, XGBoost, and CatBoost. And deep-learning algorithms used in this analysis were multilayer perceptron, Wide & Deep net-work, and residual neural net-work. Model development and internal validation were conducted using the training data, after which performance was assessed on the independent hold-out test set using ROC-AUC, accuracy, precision, recall, specificity, F1-score, and the confusion matrix. SHAP analysis was subsequently used to assess model explainability. **Abbreviations:** BMI, body mass index; MLP, multilayer perceptron; ResNet, residual neural network; ROC-AUC, area under the receiver operating characteristic curve; SHAP, SHapley Additive exPlanations; SVM, support vector machine; WHR, waist-to-hip ratio.


DATA PREPARATION
MODEL DEVELOPMENT
EVALUATION
Final Analytical Cohort (N = 22,012)
Stratified Train–Validation–Test Split
64% Training / 16% Validation / 20% Testing
Training Set (64%)
(n = 14,087)
Validation Set (16%)
(n = 3,522)
Independent Hold-out Test Set (20%)
(n = 4,403)
Data Preprocessing
• StandardScaler (numeric variables)
• OneHotEncoder (categorical variables)
• Training-data fit only
Three Feature Configurations
• BMI-only
• WHR-only
• BOTH (BMI + WHR)
Model Development
Classical Machine Learning Models
• Logistic Regression
• Linear SVM
• Random Forest
• Balanced Random Forest
• XGBoost
• CatBoost
Deep Learning Models
• Multilayer Perceptron (MLP)
• Wide & Deep Network
• Residual Neural Network (ResNet)
3-fold Stratified Cross-Validation
(On Training Set)
Internal Validation
(on Validation Set)
Model Selection
(Based on CV and Validation Performance)
Refit Selected Model on
Training + Validation Sets (80%)
(n = 17,609)
Independent Hold-out Test Set
Evaluation (20%)
Performance Metrics
• ROC-AUC
• Precision
• Specificity
• Confusion Matrix
• Accuracy
• Recall
(Sensitivity)
• F1-Score
SHAP Explainability Analysis
(On Test Set)