Research Article

American Journal of Medical Research & Reviews

Machine-Learning and Neural-Network Prediction of Diabetes Status Using Glycemic, Lipid, Renal, and Anthropometric Biomarkers: Analysis of a Public Kaggle Dataset

Ahed J Alkhatib*

Department of Legal Medicine, Toxicology and Forensic Medicine, Jordan University of Science & Technology, Jordan

***Correspondence Author:**
Ahed J Alkhatib, Department of Legal Medicine, Toxicology and Forensic Medicine, Jordan University of Science & Technology, Jordan

Received: 05 Aug 2026, **Accepted:** 21 Aug 2026, **Published:** 24 Aug 2026

Citation: Ahed J Alkhatib, Machine-Learning and Neural-Network Prediction of Diabetes Status Using Glycemic, Lipid, Renal, and Anthropometric Biomarkers: Analysis of a Public Kaggle Dataset. *American J Med Res Rev*. 2026; 1(1), 01-18.

Abstract

Background: *Diabetes mellitus is an important global metabolic disorder, and early identification remains important for reducing long-term cardiovascular, renal, and metabolic complications.*

Objective: *This study evaluated clinical, biochemical, and machine-learning markers for differentiating normal, prediabetic and diabetic status focusing on prediabetes diagnosis and models performance on exclusion of HbA1c.*

Methods: *The cleaned data set was analyzed for 1,000 participants of which 103 normal, 53 prediabetic, 844 diabetics. We used various statistics and tests to make comparisons across diabetes classes like Spearman Correlation and more. For analysis in machine-learning, repeats of clinical profiles were removed to obtain 826 records. A set of classical machine-learning models and neural networks were evaluated on binary and multiclass classification using accuracy, balanced accuracy, sensitivity/recall, F1-score, and ROC-AUC. To avoid any diagnostic circularity, we prioritized models that did not utilize HbA1c.*

Results: *After cleaning, the dataset included 103 normal, 53 prediabetic, and 844 diabetic records. After clinical deduplication, 826 unique records remained for machine-learning analysis. HbA1c and BMI showed the strongest differences across diabetes classes, with large Kruskal-Wallis effect sizes. HbA1c correlated moderately with BMI and age. In multivariable logistic regression excluding HbA1c, BMI, cholesterol, and age independently predicted diabetic status. In binary machine-learning analysis without HbA1c, random forest achieved the best performance, with 95.6% accuracy, 94.7% balanced accuracy, 96.1% sensitivity, 93.4% specificity, 97.4% F1-score, and 98.2% ROC-AUC. Neural networks performed well but did not outperform random forest; the best neural network achieved 90.4% balanced accuracy and 95.9% ROC-AUC. Multiclass classification was more difficult, mainly due to the small prediabetic group.*

Conclusion: *Information about BMI, age, cholesterol, triglycerides and TG/HDL ratio can help detect diabetes and prediabetes. Overall, the random forest models performed the best, but the identification of prediabetes was still difficult due to class imbalance and overlapping metabolic profiles.*

Keywords: Diabetes mellitus; HbA1c; BMI; lipid profile; random forest; neural network; machine learning

Introduction

Diabetes mellitus is a major global metabolic disorder and one of the most important chronic diseases affecting adult populations worldwide (Davies et al., 2026). The International Diabetes Federation Diabetes Atlas shows that diabetes is already on the increase right across the world and

this increase is projected to 2050 (International Diabetes Federation, 2025). With an increase of this level, insulin therapy will not be enough. Population screening, prevention and effective management will be essential (Duncan et al., 2026; International Diabetes Federation, 2025).

At present, three accepted diagnoses of diabetes and prediabetes are fasting plasma glucose, oral glucose tolerance testing and HbA1c (Budak et al., 2026). The reason why most diagnostic laboratories use HbA1c testing is that it reflects longer term glycemic exposure (Budak et al., 2026). It is also considered practical for application in clinical and epidemiological settings (International Expert Committee, 2009; Sacks et al., 2023; American Diabetes Association Professional Practice Committee (2026a). The preventive recommendations involve lifestyle change, reaching normal weight, and early identification in those at risk (Knowler et al., 2002; American Diabetes Association Professional Practice Committee, 2026b).

There is a major association between type 2 diabetes and age, adiposity, dyslipidemia, and overall cardiometabolism dysfunction (Putri et al., 2026). Increased body weight and obesity are well recognized to promote insulin resistance and development of type 2 diabetes (Chandrasekaran and Weiskirchen, 2024; Sulu et al., 2024). Though it should be noted, that lipid-related markers may also be clinically important (especially triglycerides, HDL cholesterol, and the triglyceride-to-HDL ratio which has been proposed to be an easily accessed surrogate marker of insulin resistance and cardiometabolic risk (Kosmas et al., 2023; Baneu et al., 2024; DeForest et al., 2024).

Over the years, machine learning has been increasingly used for diabetes prediction since it detects complex patterns in demographic, anthropometric, biochemical and clinical variables (Tasnia et al., 2026). Machine-learning models may offer good performance for type 2 diabetes prediction, however, systematic reviews call for transparent reporting, validation and careful interpretation of performance metrics (De Silva et al., 2020; Fregoso-Aparicio et al., 2021; Kiran et al., 2025). Studies involving extensive cohorts have revealed that these factors are potential predictors of diabetes, including body mass index (BMI), age, glucose-related measures, lipid markers, and other routine clinical variables (Lugner et al., 2024; Lee et al., 2025).

The present study analyzes a public Kaggle diabetes dataset containing demographic, anthropometric, renal, lipid, and glycemic biomarkers. The study combines conventional statistics, classical machine learning, and neural-network modeling to identify variables associated with diabetes status and to compare predictive performance across different modeling approaches.

Study objectives

This study aimed to evaluate diabetes-status classification using conventional statistical analysis, classical machine-learning algorithms, and neural-network models based on demographic, anthropometric, lipid, renal, and glycemic biomarkers.

Materials and Methods

Study Design and Data Source

A secondary analysis of Kaggle diabetes dataset was done for our study purpose. The data set shows medical as well as laboratory variables like age, gender, urea, creatinine, HbA1c, lipid-profile markers, BMI as well as the disease class of diabetes. Because the public dataset is de-identified, there was no patient contact involved.

Study Variables

The original dataset contains 1000 records and 14 variables. These include gender, age, urea, creatinine, HbA1c, cholesterol, triglycerides, HDL, LDL, VLDL, BMI, diabetes class. The outcome variable was CLASS, with categories normal/non –diabetic, prediabetic and diabetic.

Data Cleaning and Preprocessing

During the data cleaning phase, an investigation of class labels was conducted to trim spaces along with standardizing the gender labels. Furthermore, missing values and duplicate records were checked. Detected no missing values. A single lowercase gender code was standardized. The same patient numbers and similar clinical profiles were found. The complete cleaned dataset was applied for descriptive statistics. To minimize potential inflation of performance, clinically replicated records were excluded from machine-learning analysis. The ratio of triglycerides to HDL was calculated by dividing triglycerides by HDL. This ratio was included because prior studies described TG/HDL as an easy-to-calculate index that is linked to insulin resistance and cardiometabolic risk (Kosmas et al., 2023; Baneu et al., 2024; DeForest et al., 2024).

Conventional Statistical Analysis

Continuous variables were either described with mean and standard deviation or as median with IQR. Shapiro-Wilk test was performed to check normality. Kruskal-Wallis tests were used to compare groups of the three diabetes classes because many variables were not normally-distributed and groups sizes were unequal. Mann-Whitney U tests with Bonferroni correction were conducted for pairwise post-hoc comparisons. The Chi-square test was used to assess gender distribution across diabetes classes. Spearman correlation was carried out to determine associations of HbA1c with continuous variables. A binary logistic regression was carried out to find independent predictors of diabetic status. HbA1c was excluded from the primary logistic regression model to avoid diagnostic circularity.

Machine-Learning and Neural-Network Analysis

Machine-learning analysis was conducted using the cleaned, clinical-duplicated dataset. Two prediction tasks were assessed: one tested binary classification for diabetic versus normal/prediabetic participants and the other for multiclass

classification of normal versus prediabetic versus diabetic participants. Main predictive models excluded HbA1c due to diagnostic circularity. Secondary benchmark findings - with HbA1c's predicted diagnostic contribution.

Classical machine-learning models tested were logistic regression, random forest, extra trees, support vector machine with radial basis function kernel, gradient boosting, K-nearest neighbors, and etc. The reason for choosing these models is that they are well known linear, tree-based, kernel-based, boosting and instance based classifiers (Cover and Hart, 1967; Cortes and Vapnik, 1995; Breiman, 2001; Friedman, 2001; Geurts et al., 2006; Pedregosa et al., 2011). Standardized continuous predictors and transformed categorical variables assessed multilayer perceptron neural networks. The study evaluated three types of architectures: an architecture with one hidden layer with 8 neurons, one with one hidden layer with 16 neurons and the last one with 2 hidden layer dimensions of 16 and 8 neurons. A neural-network model was implemented based on back-propagation methods for learning internal representations (Rumelhart et al., 1986).

Since the dataset was not balanced, normal accuracy was not taken alone. We noted balanced accuracy; sensitivity; specificity; F1 score. ROC analysis is a common technique used to evaluate the discrimination of classifiers as a function of thresholds (Fawcett, 2006). Although oversampling methods, such as SMOTE, are widely used in imbalanced datasets (Chawla et al., 2002), this analysis focused on clinical deduplication, cross-validation, and class-sensitive metrics rather than synthetic record generation.

Model development, preprocessing, validation, and performance reporting were presented according to TRIPOD, TRIPOD+AI, PROBAST and PROBAST+AI (Collins et al., 2015; Wolff et al., 2019; Collins et al., 2024; Moons et al., 2025). These guidelines are designed to ensure transparent reporting and careful assessment of prediction models developed using regression or artificial-intelligence methods.

Results

Data-Quality Audit and Class Distribution

The dataset consisted of 1,000 records and 14 variables after cleaning. It has no missing values. The dataset was highly imbalanced with a majority of participants being diabetic (Table 1).

Table 1: Data-quality summary

Item	Result
Total records	1,000
Total variables	14
Missing values	0
Normal cases	103
Prediabetic cases	53
Diabetic cases	844
Female records	435
Male records	565
Duplicate patient-number rows	39
Repeated clinical profiles	174

According to Table (2) and Figure (1), the cleaned dataset consists of various classes of Diabetes. Out of 1000 individuals surveyed, 844 were insulin dependent that makes up to 84.4%. Normal participants contributed to 103 cases (10.3%), while prediabetic participants contributed to 53 cases (5.3%). The distribution shows a distinct class imbalance with the diabetic people being the majority. The small percentage of prediabetic participants may limit statistical power for prediabetes-specific comparisons and may limit predictive model performance to identify prediabetes status. As a result, follow-up analyses should interpret prediabetes-related results cautiously and adopt balanced performance metrics, rather than overall accuracy.

Table 2: Distribution of diabetes classes.

Class	Frequency (N)	Percentage (%)
Normal	103	10.3
Prediabetic	53	5.3
Diabetic	844	84.4

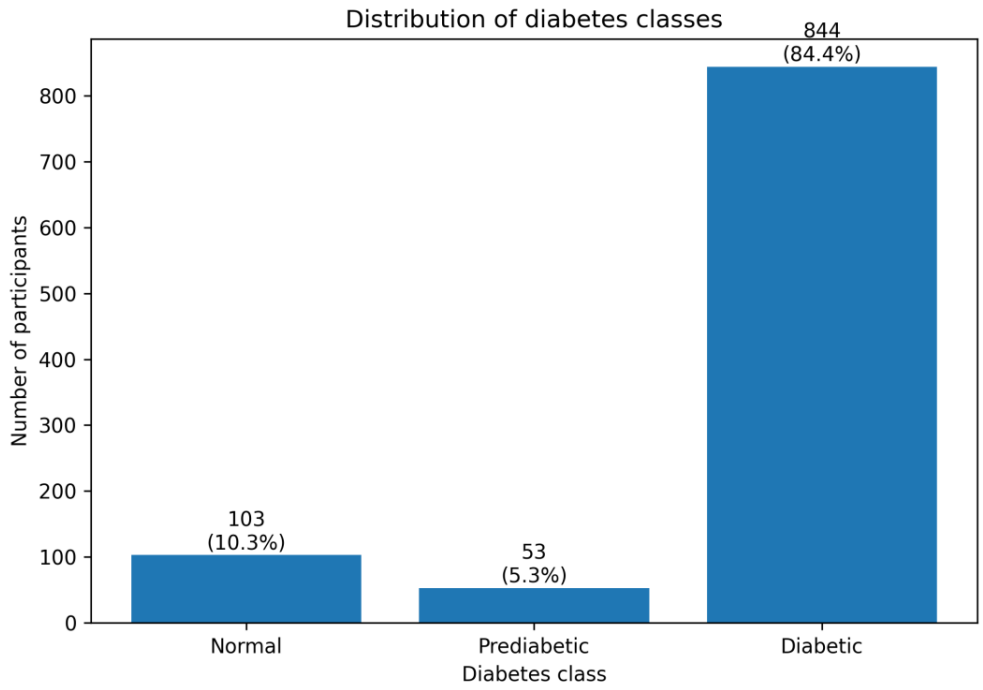

Figure 1: Distribution of diabetes classes in the cleaned dataset

Clinical and Metabolic Characteristics Across Diabetes Classes

Table (3) provides a summary of characteristics of individuals based on class of diabetes. A metabolic gradient was seen to progressively develop across normal, prediabetic, and diabetic groups. Normal and prediabetic participants had lower BMI and HbA1c values than diabetic participants, who were older. The prediabetic group's profile was in-between that of the normal and diabetic groups; this is especially true for indicators such as BMI and HbA1c. Across the classifications of diabetes, various cholesterol measures, including triglycerides (TG), VLDL, and TG/HDL, increased with class. In contrast, levels of HDL and LDL did not differ much between groups, nor did the renal markers. According to the table, HbA1c, BMI, age, triglycerides, VLDL, and TG/HDL ratio are the most informative descriptive variables for diabetes-status groups in the dataset.

Table 3: Descriptive profile by diabetes class. Values are median [IQR].

Variable	Normal	Prediabetic	Diabetic
Age	44.00 [38.50-50.00]	48.00 [35.00-50.00]	55.00 [53.00-60.00]
BMI	22.00 [21.00-24.00]	24.00 [23.00-25.00]	30.00 [28.00-33.00]
HbA1c	4.90 [4.10-5.10]	6.00 [5.90-6.10]	8.80 [7.20-10.40]
Cholesterol	4.20 [3.65-4.80]	4.70 [4.00-5.30]	4.90 [4.10-5.70]
Triglycerides	1.30 [1.00-1.80]	1.80 [1.30-2.40]	2.10 [1.50-3.00]
HDL	1.10 [0.90-1.30]	1.00 [0.80-1.40]	1.10 [0.90-1.30]
LDL	2.60 [1.90-3.25]	2.50 [1.90-3.20]	2.50 [1.80-3.30]
VLDL	0.70 [0.55-1.00]	0.80 [0.60-1.30]	1.00 [0.70-1.50]
TG/HDL ratio	1.25 [0.90-1.62]	1.50 [1.21-2.75]	1.91 [1.25-3.00]
Urea	4.40 [3.30-5.45]	4.40 [3.40-5.00]	4.60 [3.70-5.80]
Creatinine	55.00 [45.50-71.50]	59.00 [53.00-64.00]	60.50 [48.00-73.00]

Explanation of the BMI and HbA1c boxplots

As shown in Figures (2 and 3), depicted distributions of BMI and HbA1c show marked metabolic differences between diabetes classes. As we move from normal to prediabetic to diabetic participants, the BMI increased progressively in an upward direction, with the median BMI of the diabetic group the highest and also the most variable. This trend indicates that a higher level of fatness worsens diabetes. HbA1c was also significantly different among the three groups. HbA1c values for normal participants were the lowest, while those for prediabetic participants were tightly clustered around intermediate values; the diabetic participants, on the other

hand, exhibited significantly higher values that were more variable. All in all, the boxplots back the description, stating that BMI and HbA1c are key distinguishing features for normal, prediabetic, and diabetic participants in the dataset.

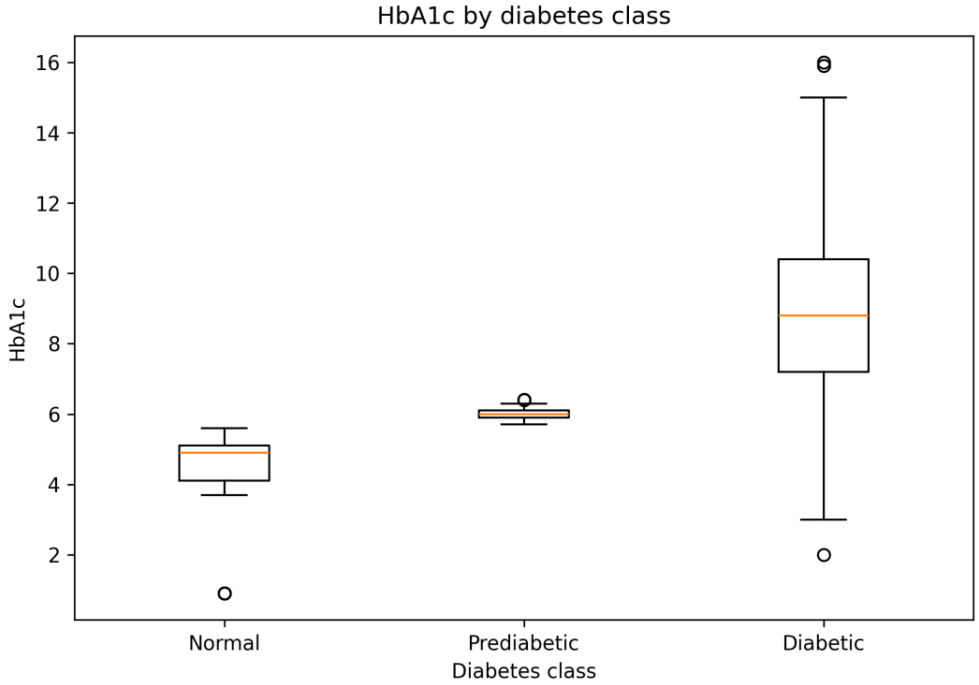

Figure 2: HbA1c distribution across normal, prediabetic, and diabetic groups.

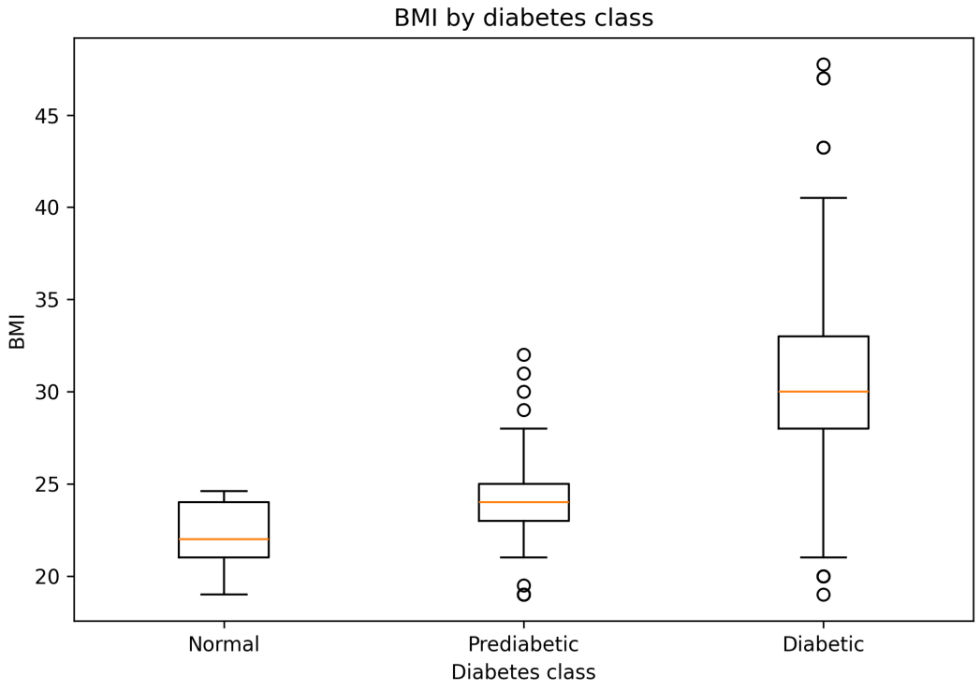

Figure 3: BMI distribution across normal, prediabetic, and diabetic groups.

Non-parametric Comparison of Clinical and Metabolic Variables Across Diabetes Classes

Table (4) represents Kruskal-Wallis Analyses of Clinical as well as Biochemical Variables of Normal, Prediabetic and Diabetic Groups. There were highly significant differences in HbA1c, BMI, age, triglycerides, VLDL, TG/HDL ratio, cholesterol and urea. As per effect size, HbA1c and BMI were best in discriminating between diabetic and non-diabetic individuals with similar dietary habits. Groups exhibited significantly different ages with a moderate effect size. Lipid markers, namely triglycerides, VLDL, TG/HDL ratio, cholesterol, and others, showed statistically

significant differences; however, the effect size was small. Urea was statistically significant but paltry effect size makes it practically unimportant. There were no significant differences in creatinine, HDL, and LDL levels across diabetes classes. According to the analysis, the most identifiable variables between the diabetes-status groups were HbA1c, BMI and age whereas the lipid-related markers gave some additional but marginally identifiable information.

Table 4: Kruskal-Wallis comparisons across diabetes classes.

Variable	H statistic	p-value	Effect size	Interpretation
HbA1c	334.05	<0.001	0.333	Large
BMI	324.85	<0.001	0.324	Large
Age	230.88	<0.001	0.230	Moderate
Triglycerides	62.04	<0.001	0.060	Small
VLDL	54.83	<0.001	0.053	Small
TG/HDL ratio	44.65	<0.001	0.043	Small
Cholesterol	35.57	<0.001	0.034	Small
Urea	8.25	0.016	0.006	Negligible
Creatinine	2.32	0.314	0.000	Not significant
HDL	1.21	0.546	0.000	Not significant
LDL	0.35	0.841	0.000	Not significant

Post-hoc Pairwise Comparisons Across Diabetes Classes

Normal, prediabetic and diabetic participants significantly differed in all the post-hoc pairwise comparisons (Table 5). HbA1c and BMI were examined in this analysis. There was a significant difference between all paired comparison of HbA1c and BMI. There was a progressive rise from normal status to prediabetes, and from prediabetes to diabetes. The age between participants with diabetes with normoglycemics and prediabetics was significantly different. Thus, the diabetics were older and the normal and prediabetes were not distinct in age. Different lipid-related markers showed selective differences. The study demonstrated that cholesterol, triglycerides and TG/HDL ratio differed significantly between normal and prediabetic or diabetic participants. This suggests that lipid abnormalities may be present during the prediabetic stage. The major differences in VLDL were in diabetic comparisons. In conclusion, the current post-hoc analysis lends support to the significant association between HbA1c and BMI with various stages of diabetes. Further analysis reveals that lipid markers add weight to the evidence of early cardiometabolic disturbance.

Table 5: Significant post-hoc differences.

Variable	Significant pairwise differences
Age	Normal vs diabetic; prediabetic vs diabetic
HbA1c	Normal vs prediabetic; normal vs diabetic; prediabetic vs diabetic
BMI	Normal vs prediabetic; normal vs diabetic; prediabetic vs diabetic
Cholesterol	Normal vs prediabetic; normal vs diabetic
Triglycerides	Normal vs prediabetic; normal vs diabetic
VLDL	Normal vs diabetic; prediabetic vs diabetic
TG/HDL ratio	Normal vs prediabetic; normal vs diabetic

Association Between Gender and Diabetes Class

Table (6) displays the Chi-square analysis of the relationship between gender and class of diabetes. The test results indicated that a statistically significant association between gender and diabetes status. Specifically, a result of $\chi^2 = 18.20$, $p < 0.001$ indicated that the distribution of males and females varied across normal, prediabetic, and diabetic groups. The effect of the treatment was weak (Cramer's V = 0.135). This indicates that gender was not meaningfully associated with diabetes class despite being statistically so. Thus, gender should not be considered as an important variable in this dataset.

Table 6: Chi-square analysis of gender and diabetes class

Test	Chi-square	df	p-value	Cramer's V
Gender x diabetes class	18.20	2	<0.001	0.135

Correlation of HbA1c with Anthropometric, Lipid, and Renal Markers

According to table 7, a Spearman correlation analysis was conducted on the parameters of HbA1c. A moderate positive correlations was reported between HbA1c and BMI as well as age. Indicating a higher glycemic burden was associated with higher adiposity and older age. The lipid-related markers like triglycerides, VLDL, TG/HDL ratio and cholesterol showed weak but significant positive correlation with HbA1c levels indicating that there is a mild association between worsening glycemic status and associated dyslipidemic changes. A statistically significant but negligible negative correlation was, however, found between creatinine and HbA1c which may not be meaningful clinically. No significant relationship was found between HbA1c and urea, HDL, or LDL. In general, the HbA1c was most closely associated with BMI and age, while the cholesterol markers were the next closest evidence of metabolic disturbance.

Table 7: Spearman correlations with HbA1c

Variable	Spearman rho	p-value	Strength
BMI	0.417	<0.001	Moderate
Age	0.411	<0.001	Moderate
Triglycerides	0.232	<0.001	Weak
VLDL	0.222	<0.001	Weak
TG/HDL ratio	0.193	<0.001	Weak
Cholesterol	0.152	<0.001	Weak
Creatinine	-0.082	0.009	Negligible
Urea	0.015	0.627	Not significant
HDL	-0.027	0.390	Not significant
LDL	-0.023	0.474	Not significant

As evidenced by the Spearman correlation heatmap (Figure 4), HbA1c is most highly correlated with BMI and age, indicating that a greater glycemic burden is likely to occur in older participants and those with more adiposity. All the individual characteristics of HbA1c showed weak positive correlations with triglycerides, VLDL, TG/HDL ratio, cholesterol with a mild relationship with worsening glycemic status with varying dyslipidemic markers. Higher associations were detected between lipid variables, notably the triglyceride and TG/HDL ratio and the triglyceride and VLDL pair. It was expected that HDL and the TG/HDL ratio would show a negative correlation as HDL is in the denominator of this calculated index. Urea and creatinine also had a moderately positive correlation suggesting clustering associated with renal function. In conclusion, the heatmap supports the findings that BMI and age are likely the two main non-glycemic correlate of HbA1c while the lipid variables give an insight into the overlapping metabolic information that is of additional value.

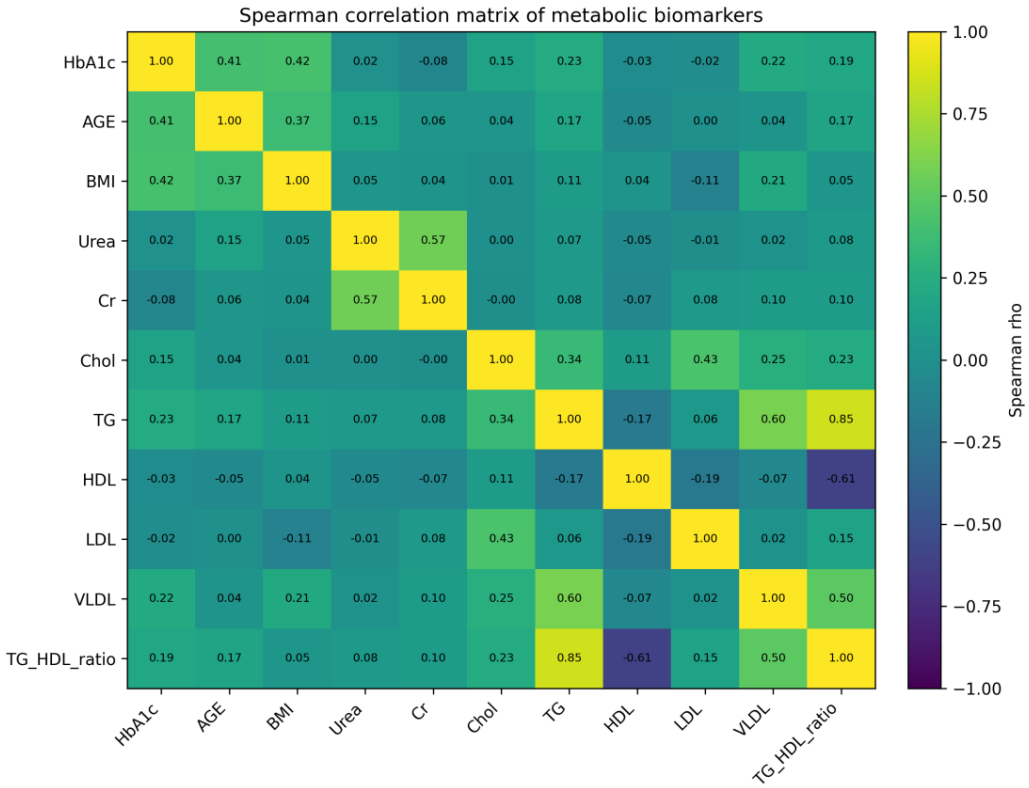

Figure 4: Correlation heatmap of glycemic, lipid, renal, and anthropometric markers

Multivariable Logistic Regression Analysis of Non-Glycemic Predictors of Diabetes

Multivariate logistic regression analysis of diabetic status (HbA1c excluded to avoid diagnostic circularity) shown in Table (8). The analysis indicates that diabetic status is significantly predicted by BMI, cholesterol and age. Each unit increase in BMI was found to be linked with almost double odds of having diabetes. Cholesterol was also independently associated with diabetic status while age was positively and less significantly associated. After adjustment, the trend for triglycerides was positive but not statistically significant. Urea, creatinine, male sex, HDL and LDL weren't significant independent predictors. The findings of the regression analysis overall suggest that the non-glycemic indicators particularly BMI, cholesterol and age had a considerable bearing in differentiating the diabetic participants from the normal and the pre-diabetic participants in the given dataset.

Table 8: Multivariable logistic regression predicting diabetic status. Outcome: diabetic versus normal/prediabetic; HbA1c excluded

Predictor	Adjusted OR	95% CI	p-value
BMI	2.00	1.75-2.28	<0.001
Cholesterol	2.38	1.70-3.31	<0.001
Age	1.07	1.03-1.10	<0.001
Triglycerides	1.29	0.96-1.73	0.086
Urea	1.06	0.89-1.27	0.494
Creatinine	1.01	0.99-1.02	0.437
Male gender	0.99	0.55-1.79	0.974
HDL	1.00	0.53-1.88	0.993
LDL	0.93	0.65-1.33	0.693

Adjusted Odds Ratios for Non-Glycemic Predictors of Diabetes

As shown in Figure (5), the forest plot shows odds ratios from a multivariable logistic regression model predicting diabetic status after the excluding of HbA1c. The only factors to predict diabetic status with all their confidence intervals above the null value of 1 were cholesterol, BMI and age. The most powerful effects were exhibited by cholesterol and BMI, while age had a weaker but consistently positive association. The triglycerides trend was positive but not

statistically significant as the confidence interval crossed 1. Urea, creatinine, HDL, LDL and being a male gender were not significant independent predictors. In general, the figure indicates that the association with being diabetic in this dataset primarily pertains to adiposity, cholesterol and age, when HbA1c is excluded in the model.

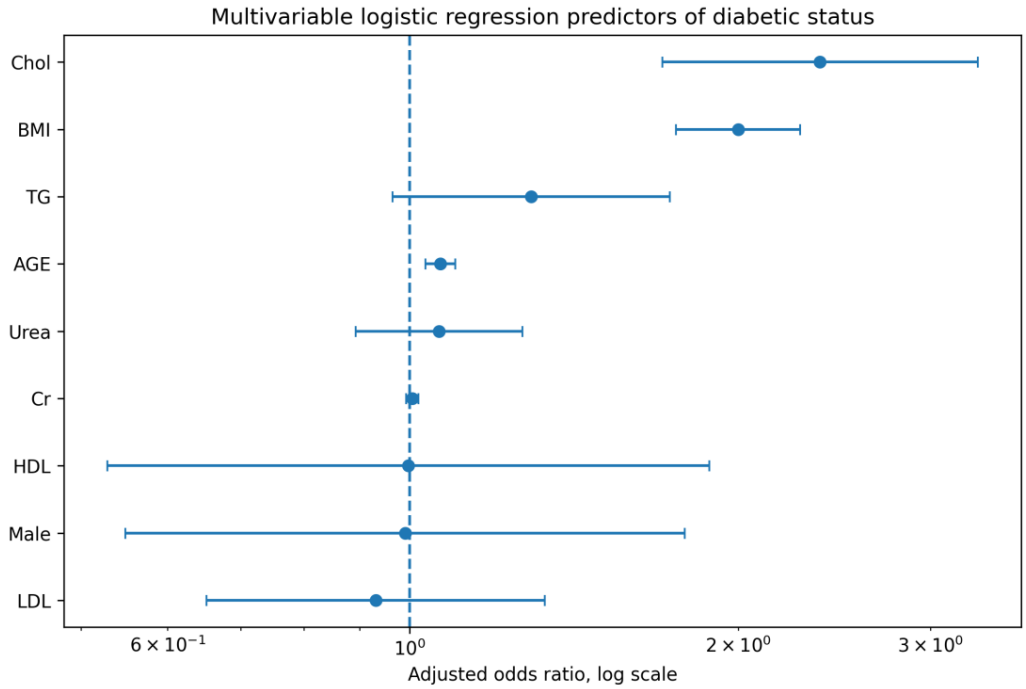

Figure 5: Adjusted odds ratios for predictors of diabetic status

Machine-Learning Results

Following the deletion of repetitive clinical profiles, the unique 826 record was retained (Table 9). The prediction dataset had 96 normal records, 40 prediabetic records, and 690 diabetic records. While deduplication decreased the likelihood of overlooked overestimations due to repeating profiles, we still can have high imbalance in our data set with diabetic cases as the major class and prediabetic cases as the smaller class. The discrepancy would have serious implications for model evaluation since overall accuracy would exaggerate model performance due to favoring the majority diabetic class. Hence, accuracy, recall, specificity, F1 score, macro F1 score, and ROC-AUC were used to evaluate models. The reduced prevalence of records for prediabetes further explains the observed increased difficulty in multiclass, prediabetes prediction.

Table 9: Dataset used for machine-learning analysis.

Class	N
Normal	96
Prediabetic	40
Diabetic	690
Total	826

Binary Machine-Learning Prediction of Diabetic Status Without HbA1c

Table (10) and Figure (6) present the findings related to the cross-validated performances of binary machine-learning models which attempt to distinguish between diabetics and normal and prediabetic participants without the addition of HbA1c. The random forest model obtained an accuracy of 95.6%, a balanced accuracy of 94.7%, a sensitivity of 96.1%, 93.4% specificity, and 98.2% AUC-ROC values for validation data. According to the ROC-AUC, the greater sensitivity of gradient boosting compared to that of logistic regression and a relatively lower specificity of the former holds that it is more likely to classify a non-diabetic record as diabetic. Other classifiers like Extra trees, SVM-RBF and logistic regression also manage to discriminate well. K-Nearest Neighbors produced high sensitivity, but comparatively lower specificity and balanced accuracy. So, on an imbalanced dataset, overall accuracy may give misleading result. In general, the random forest did the best job of classifying the diabetic and normal or prediabetic records.

Table 10: Binary machine-learning performance without HbA1c. Outcome: diabetic versus normal/prediabetic.

Model	Accuracy %	Balanced accuracy %	Sensitivity %	Specificity %	F1-score %	ROC-AUC %
Random forest	95.6	94.7	96.1	93.4	97.4	98.2
Logistic regression	90.8	91.3	90.6	91.9	94.2	95.9
Extra trees	92.2	90.6	93.0	88.2	95.3	97.8
SVM-RBF	91.0	90.5	91.3	89.7	94.4	96.4
Gradient boosting	94.9	90.5	97.1	83.9	97.0	98.4
KNN	92.0	83.1	96.4	69.9	95.3	95.2

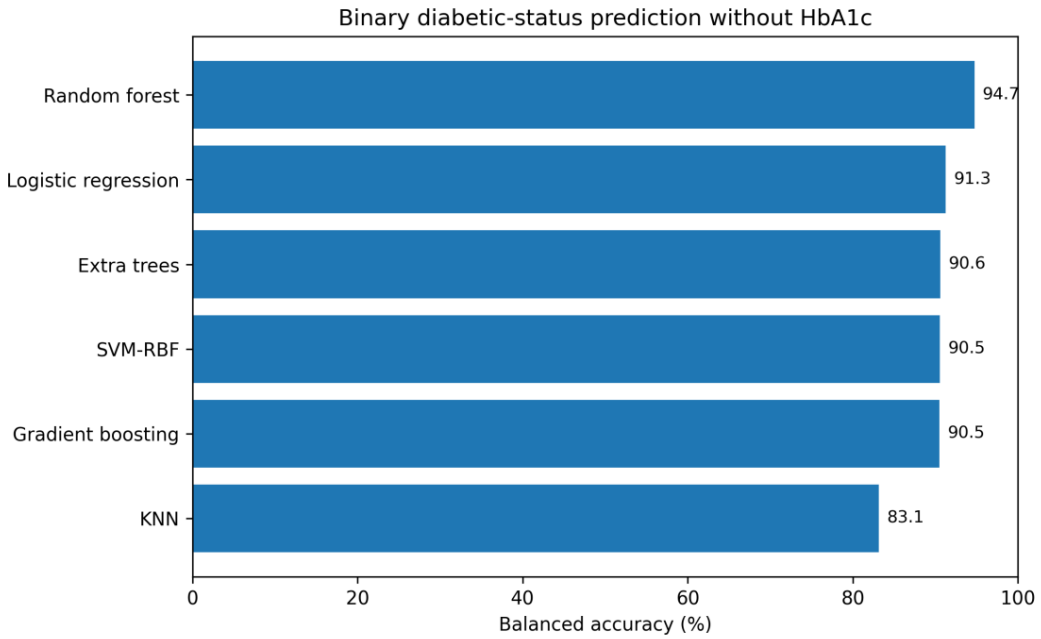

Figure 6: Binary machine-learning balanced accuracy without HbA1c

ROC Curve for Binary Diabetic-Status Prediction with and Without HbA1c

Both models exhibit excellent binary discrimination, as shown by the ROC curves (Figure 7). The AUC for gradient boosting with HbA1c and random forest without HbA1c were 1.000 and 0.985, respectively. It was anticipated that the HbA1c-inclusive model would fare excellently because this factor contributed directly to diabetes classification and may therefore introduce diagnostic circularity. In addition, the random-forest model showed excellent discrimination following the exclusion of HbA1c, indicating that non-glycemic clinical and biochemical variables can separate diabetic and normal/prediabetic records. However, external validation is necessary prior to clinical use.

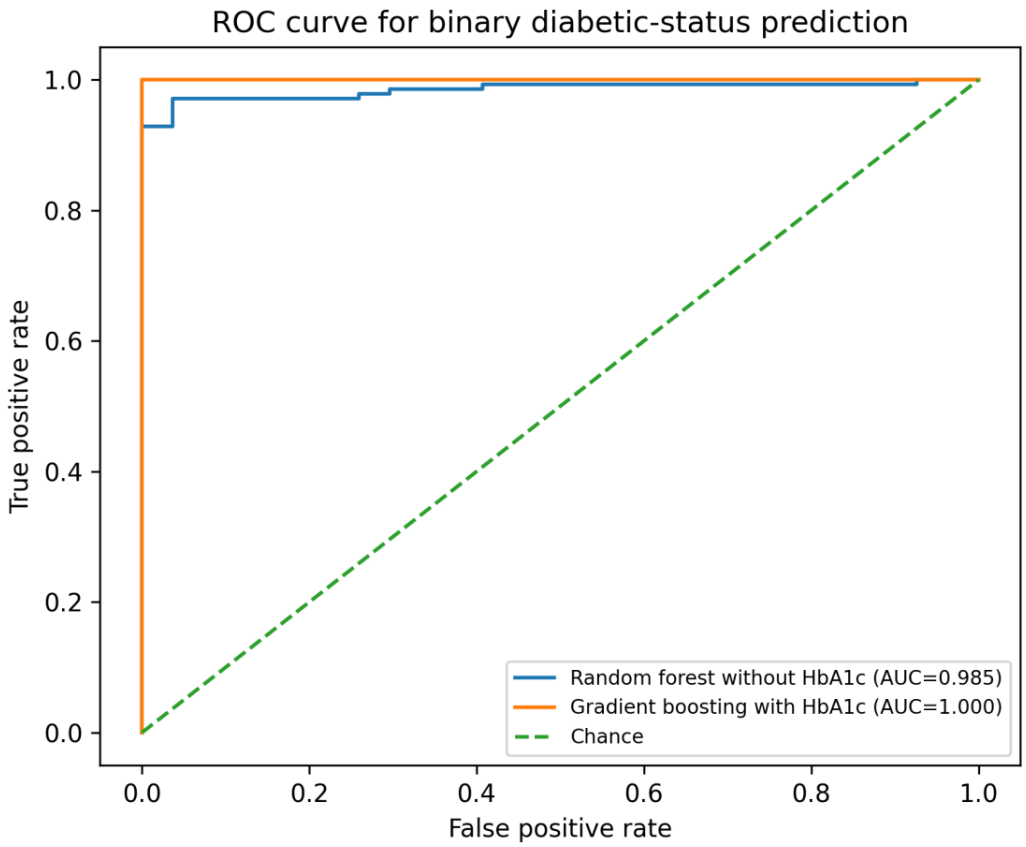

Figure 7: ROC curve for the best binary machine-learning model

Benchmark Binary Machine-Learning Performance with HbA1c

Inclusion of HbA1c as a predictor in binary ML models improves benchmark performance as shown in Table (11). Gradient boosting gave the best performance with an accuracy of 98.5% and a balanced accuracy of 97.6%. Its sensitivity, specificity, and ROC-AUC equal to 99.0%, 96.3%, and 99.8%, respectively. Random forest and extra trees also achieved excellent discrimination reaching ROC AUC values of 99.4% and 98.9%, respectively. When evaluated, the performance continued to be higher than that of the models excluding HbA1c. This confirmed that HbA1c had a large discriminatory contribution. Because HbA1c was used to define diabetes status, these estimates are vulnerable to diagnostic circularity, and should be viewed benchmark results Consequently, the HbA1c-excluded analysis is a more informative assessment of prediction through non-glycemic characteristics.

The AUC of 1.000 for the ROC figure should be considered as a performance on the test set, while the 99.8% should be identified as the mean cross-validated AUC.

Table 11: Binary benchmark machine-learning performance with HbA1c.

Model	Accuracy %	Balanced accuracy %	Sensitivity %	Specificity %	F1-score %	ROC-AUC %
Gradient boosting	98.5	97.6	99.0	96.3	99.1	99.8
Random forest	96.5	95.5	97.0	94.1	97.9	99.4
Extra trees	94.3	94.5	94.2	94.8	96.5	98.9

Multiclass Prediction of Diabetes Status Without HbA1c

Table (12) shows multiclass performance of machine learning models which classify normal, prediabetic and diabetic records without HbA1c. Out of all the classifiers chosen in the present analysis, random forest attained the best overall results with an accuracy of 92.6%, balanced accuracy of 76.3%, macro F1-score of 75.1%, and one versus rest ROC-AUC of 96.5%. The extra trees also show strong discrimination while the logistic regression achieves similar balanced accuracy despite lower accuracy overall. Gradient boosting and KNN yield high accuracy but substantial low balanced accuracy, which indicates high performance for the majority diabetic class but poor recognition of the other minor classes. In general, random forest produced the most balanced multiclass performance. Nonetheless, the balanced accuracy and macro F1-score indicate distinguishing

prediabetes is difficult because of its small sample size and overlapping non-HbA1c metabolic profile. Thus class-wise recall and confusion matrices should accompany these aggregate measures.

Table 12: Multiclass machine-learning performance without HbA1c. Outcome: normal versus prediabetic versus diabetic.

Model	Accuracy %	Balanced accuracy %	Macro F1-score %	ROC-AUC OvR %
Random forest	92.6	76.3	75.1	96.5
Logistic regression	85.8	74.0	66.1	92.8
Extra trees	90.4	74.0	71.9	96.2
SVM-RBF	86.4	72.7	66.2	94.4
Gradient boosting	92.1	68.5	69.5	95.5
KNN	90.7	62.9	66.8	93.4

Permutation Feature Importance for Random-Forest Classification Without HbA1c

According to the permutation feature-importance analysis, the dominant predictor in the random-forest model without HbA1c was BMI. Using a randomization paradigm for BMI generates a drop in balanced accuracy of about 42 percentage points, which is much greater than any of the other variables. Cholesterol was the second key predictor, with lesser but still important contributions from triglycerides, age, and VLDL. TG/HDL ratio and LDL contributed minimally to the information, while creatinine, urea, sex, and HDL were almost unimportant. The model primarily relied on adiposity as a non-glycemic factor to differentiate amongst diabetes-status classes (Figure 8).

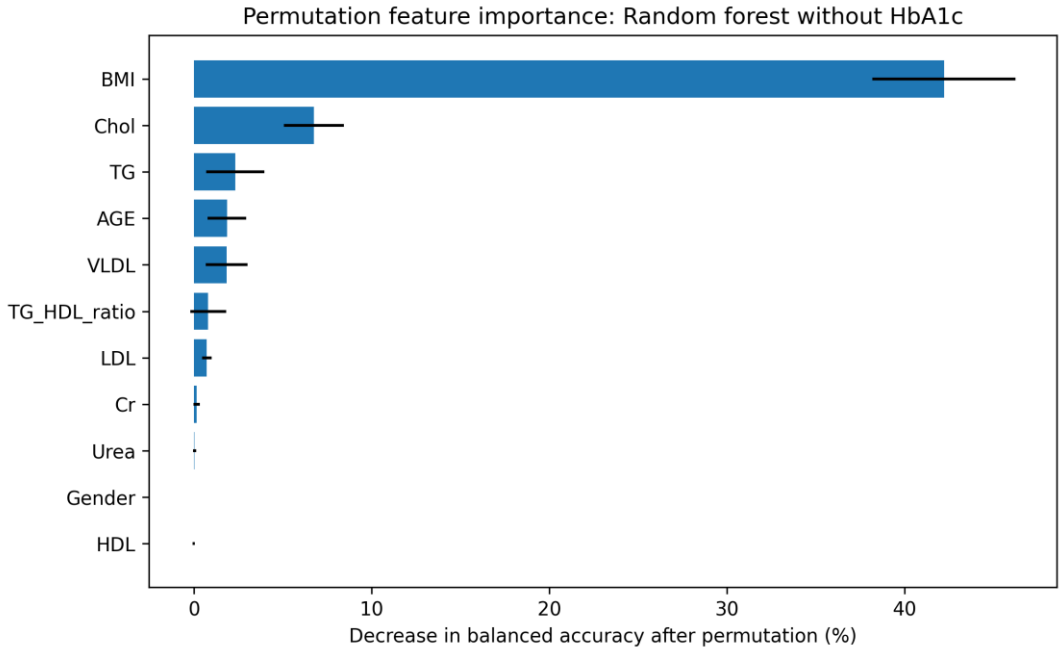

Figure 8: Feature importance of the best random forest model without HbA1c.

Neural-Network Results

Binary Neural-Network Prediction of Diabetic Status Without HbA1c

Table (13) presents the findings for the binary neural-network performance to predict diabetic status without HbA1c levels. The model NN-1, with one hidden layer having 8 neurons, achieved the highest balanced accuracy, highest specificity, and highest ROC-AUC indicating the best overall discrimination balance. NN-3 demonstrated superior accuracy, sensitivity, and F1-score but lower specificity indicating it recognizes diabetic cases at the cost of minority-class classification. NN-2 showed moderate capability. The prediction accuracy did not consistently improve by adding network complexity, nor did any of the neural-network models exceed the performance of the random-forest classifier. In general, the results suggested that a fairly

simple neural architecture was sufficient and that tree-based ensembles worked better for this data.

Table 13. Binary neural-network performance without HbA1c.

Model	Architecture	Accuracy %	Balanced accuracy %	Sensitivity %	Specificity %	F1-score %	ROC-AUC %
NN-1	8 neurons	88.4	90.4	87.4	93.4	92.6	95.9
NN-3	16-8 neurons	90.3	89.8	90.6	88.9	94.0	95.1
NN-2	16 neurons	89.0	89.6	88.7	90.4	93.0	95.4

Binary Neural-Network Benchmark Performance with HbA1c

The performance of the present binary neural-network benchmark, incorporating HbA1c, is presented in Table (14). NN-2 reaches the strongest results overall at 91.0% accuracy, 92.6% balanced accuracy, 94.4% F1-score and 97.3% ROC-AUC. The NN-1 showed the same balanced accuracy and the highest specificity. Its lower sensitivity indicates it misses more diabetic cases. NN-3 showed a more balanced sensitivity and specificity than previous architectures, but did not outperform the single-layer architectures which are simpler than NN-3. Including HbA1c improved the neural-network discrimination modestly compared to the HbA1c-excluded analysis, while tree-based ensemble models remained superior. The results should be construed as benchmark estimates as HbA1c was not modified in the definition of the outcome.

Table 14: Binary neural-network benchmark with HbA1c

Model	Architecture	Accuracy %	Balanced accuracy %	Sensitivity %	Specificity %	F1-score %	ROC-AUC %
NN-1	8 neurons	89.1	92.6	87.4	97.8	93.0	97.0
NN-2	16 neurons	91.0	92.6	90.3	94.8	94.4	97.3
NN-3	16-8 neurons	90.4	90.4	90.4	90.4	93.9	95.7

Multiclass Neural-Network Prediction Without HbA1c

Table (15) presents the multiclass performance of neural-network models for classifying normal, prediabetic, and diabetic records without HbA1c. NN-3 achieved the strongest results across all metrics, including 85.1% accuracy, 68.5% balanced accuracy, a 62.6% macro F1-score, and a one-versus-rest ROC-AUC of 92.3%. Nevertheless, the marked difference between its macro and weighted F1-scores indicates that performance was substantially stronger for the majority diabetic class than for the minority classes. NN-1 showed slightly better balanced accuracy than NN-2, despite lower overall accuracy. Compared with the classical models in Table (12), all neural networks performed less effectively, particularly in balanced accuracy and macro F1-score. These findings indicate that multiclass discrimination, especially recognition of prediabetes, remained challenging without HbA1c.

Table 15: Multiclass neural-network performance without HbA1c

Model	Architecture	Accuracy %	Balanced accuracy %	Macro F1-score %	Weighted F1-score %	ROC-AUC OvR %
NN-3	16-8 neurons	85.1	68.5	62.6	87.0	92.3
NN-1	8 neurons	79.9	67.6	58.5	82.9	90.6
NN-2	16 neurons	80.8	62.8	56.7	83.7	90.9

Comparison of Binary Balanced Accuracy Between Classical Machine-Learning and Neural-Network Models Without HbA1c

The random forest has obtained the highest balanced accuracy of 94.7% for binary diabetic-status prediction without HbA1c. With an accuracy of 91.3 percent, Logistic Regression was second-best, while extra trees, SVM-RBF, gradient boosting, and the three neural-network architectures were all fairly similar at about 89.6% to 90.6%. Out of all the neural networks, the simplest eight-neuron architecture gave an optimal Balance Accuracy. KNN performed the worst with a balanced accuracy of 83.1%. All in all, neural networks were competitive in discrimination but not better than the best classical models, random forest in particular, suggesting tree ensemble learning was a better fit for the tabular data (Figure 9).

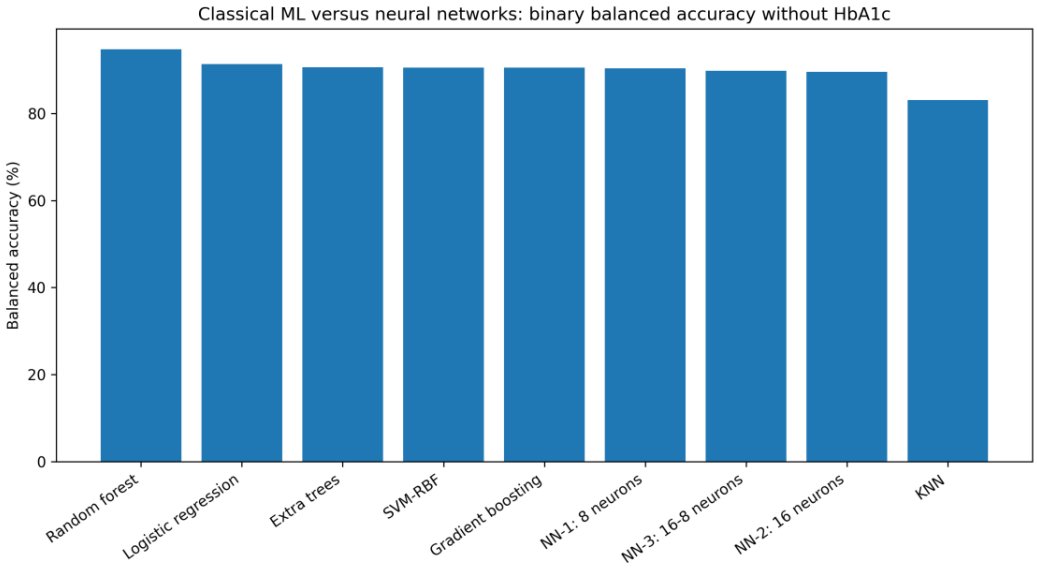

Figure 9: Classical machine-learning versus neural-network balanced accuracy for binary classification.

Training-Loss Curve for the Eight-Neuron Binary Neural Network

The training-loss curve for the eight-neuron binary neural network showed a smooth decline across the training epochs. Loss decreased rapidly during the early stages of training and then declined more gradually, reaching approximately 0.19 by the final epoch. The stabilization of the curve after approximately 30–40 epochs suggests that the model approached convergence, with only limited improvement from additional epochs. The absence of marked fluctuations indicates stable optimization. Nevertheless, because only training loss is displayed, the curve cannot independently exclude overfitting; comparison with validation loss would be required to evaluate model generalization.

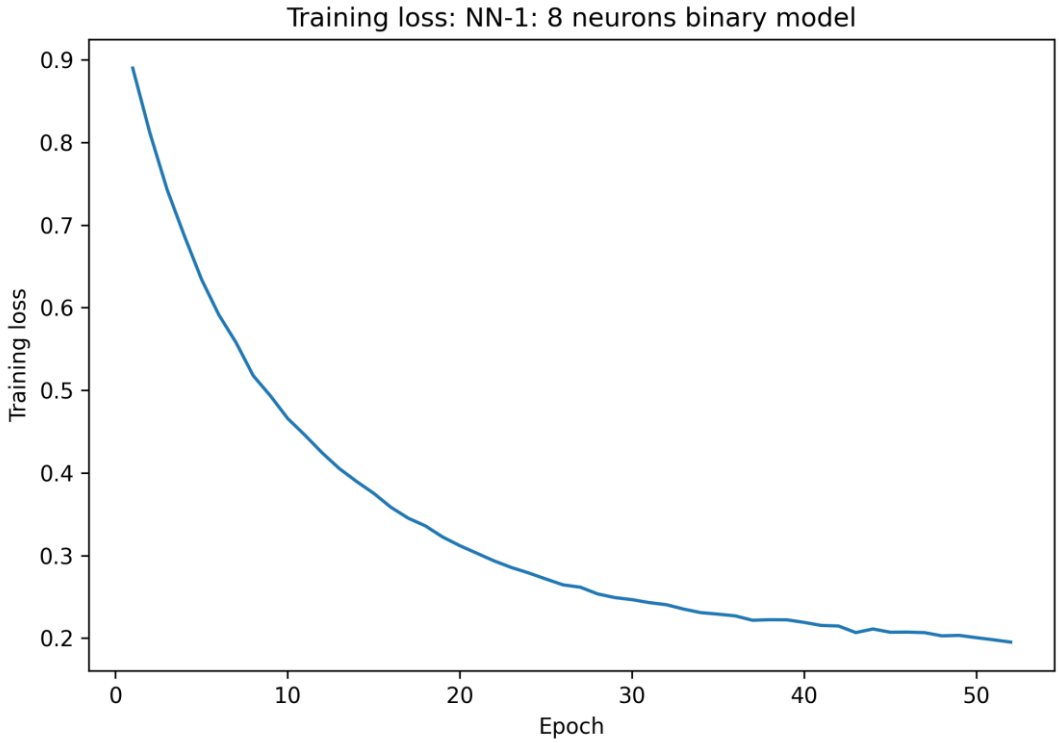

Figure 10: Training-loss curve for the best binary neural-network model.

Summary of the Best-Performing Prediction Models

Table (16) presents the optimal machine learning task model per task. The random forest was the strongest model for binary and multiclass predictions without HbA1c with a balanced accuracy of 94.7% and 76.3%, respectively. The tree-based ensemble learning was well suited to the structured data of the competition, as it regularly outperformed the corresponding neural-network models. The performance

with multiple classes was lower than binary classes due to the challenge the small prediabetic class posed on non-HbA1c variables. According to the benchmark performance of Gradient boosting, the inclusion of HbA1c resulted in the highest performance, but this should be interpreted carefully as it directly contributed to defining the outcome. In general, the most informative and clinically relevant predictive performance was of random forest without HbA1c.

Table 16: Best classical machine-learning and neural-network models.

Task	Best model	Balanced accuracy %	ROC-AUC %
Binary without HbA1c	Random forest	94.7	98.2
Binary neural network without HbA1c	NN-1	90.4	95.9
Multiclass without HbA1c	Random forest	76.3	96.5
Multiclass neural network without HbA1c	NN-3	68.5	92.3
Binary benchmark with HbA1c	Gradient boosting	97.6	99.8

Discussion

The study combines traditional statistics and neural-network modeling using machine learning to analyze the diabetes status of a public dataset which contains demographic, anthropometric, renal, lipid, and glycemic biomarkers. Diabetes class had strong correlations with HbA1c, BMI, and disease duration in this study. The markers related to lipids triglyceride, VLDL, cholesterol and TG/HDL ratio were also different across the groups although with smaller effect sizes.

HbA1c consistently increased across the normal, prediabetic, and diabetic classes with the most conventional finding. Due to this, HbA1c is directly used as a diabetes diagnosis measure, and this expected because HbA1c is directly used for diabetes diagnoses (Sacks et al., 2023; International Expert Committee, 2009; American Diabetes Association Professional Practice Committee, 2026a).

In this study, BMI stood out as the strongest non-glycemic marker. The median BMI in normal participants was 22.0 while that in diabetic participants was 30.0 and logistic regression shows that BMI predicted diabetic status independently. The evidence indicates that adiposity and weight gain are primary causes of insulin resistance and type 2 diabetes and impact the potential for diabetes remission through obesity management. (Knowler et al., 2002; Chandrasekaran and Weiskirchen, 2024; Sulu et al., 2024).

The lipid results also support the metabolic interpretation of the data. There were differences in the Triglycerides, VLDL, cholesterol, and TG/HDL ratio across various diabetes groups. The TG/HDL ratio was previously reported as a practical marker related to insulin resistance as well as cardiometabolic risk. However, note that interpretations were population-dependent (Kosmas et al., 2023; Baneu et al., 2024; DeForest et al., 2024;).

The random forest yielded the highest classification performance among the binary classifiers without-HbA1c. This result seems reasonable, as tree-based models can capture nonlinear associations and variable interactions without requiring strict distributional assumptions (Breiman, 2001; Geurts et al., 2006). The random forest and the extra trees performance shows a strong match with the systematic reviews showing that the tree-based and the assembling techniques are often very competitive in the diabetes prediction studies (De Silva et al., 2020; Fregoso-Aparicio et al., 2021; Kiran et al., 2025).

The logistic regression also performed reasonably satisfactorily suggesting that part of the dataset pattern was linearly separable. Support vector machine and gradient boosting also performed well, which was to be expected as both techniques are widely used in supervised classification problems (Cortes and Vapnik, 1995; Friedman, 2001). The worse balanced accuracy performance observed for k-nearest neighbors may reflect sensitivity to scaling, class imbalance, and/or local neighborhood structure (Cover and Hart, 1967).

When HbA1c was included, the performance of the models improved a lot. However, since HbA1c is diagnostically relevant for classification of diabetes, models which include HbA1c may partly reproduce the diagnostic label instead of providing independent prediction. As such, the non-HbA1c models are more useful for discerning whether other metabolic and anthropometric markers can classify diabetes status.

Neural-network models did a good job but did not beat random forest. A single hidden layer with eight neurons worked best as it suggests that deeper architectures were not required for this dataset. The results aligned with the common behavior present in tabular biomedical datasets of small-to-medium size where tree-based methods tend to perform well while neural networks may not offer an additional advantage unless larger or more complex structure data available (Rumelhart et al., 1986; Fregoso-Aparicio et al., 2021; Kiran et al., 2025).

Binary prediction was easier than multiclass prediction. The small size of the prediabetic group was the main challenge affecting balanced accuracy macro F. It shows that ordinary accuracy is limited in imbalanced datasets. It supports the use of balanced accuracy, class-specific metrics, and ROC-

AUC than accuracy alone (Chawla et al., 2002; Fawcett, 2006). The results should be interpreted in light of reporting quality and risk of bias. TRIPOD and TRIPOD+AI highlight transparent reporting of prediction-model development and evaluation, while PROBAST and PROBAST+AI emphasize risk-of-bias and applicability assessment (Collins et al., 2015a, b; Wolff et al., 2019; Collins et al., 2024; Moons et al., 2025). Consequently, despite the high internal performance of present models, external validation is required before clinical use.

According to recent research, body mass index (BMI), age, glycemic markers (plasma glucose levels), and lipid-related indicators (cholesterol and triglyceride levels) are among the most important predictors of future type 2 diabetes (Lugner et al., 2024; Lee et al., 2025). The current findings accord with this pattern, but use of publicly available. secondary data limits our causal inference and clinical generalizability.

Strengths

There are several strengths of this study. First, traditional statistical analysis was combined with classical machine learning and neural-network approaches. To minimize diagnostic circularity, we excluded HbA1C from the primary predictive models. The third part stresses more on the balanced accuracy, sensitivity, specificity, F1 score, and ROC-AUC rather than accuracy. Fourth, duplicate clinical profiles were removed before machine learning analysis to avoid possible performance inflation.

Limitations

This study has restrictions. To begin with, the dataset is public and secondary, hence the original sampling method and clinical context are limited. Additionally, the diabetic group was much larger than the normal and prediabetic group, creating class imbalance. It was not possible to check external validity from an independent clinical dataset. The fourth omission concerned a variety of unavailability such as fasting glucose, fasting insulin, medication history, diabetes duration, blood pressure, family history, and lifestyle variables. Insulin resistance indices, such as HOMA-IR, could not be calculated. Moreover, the inclusion of HbA1c (or glucose) may be justified as other diabetes complications models have inclusion of similar measures.

Conclusion

This study demonstrated that diabetes status in a public Kaggle dataset was strongly associated with HbA1c, BMI, age, and lipid-related markers. Conventional statistics showed clear metabolic separation across normal, prediabetic, and diabetic groups. In predictive modeling, random forest was the strongest model, outperforming neural-network models in both binary and multiclass classification. Neural networks performed acceptably but did not provide additional advantage over tree-based models. The findings suggest that BMI, age, and lipid-profile markers can support diabetes-status classification, but external validation using independent clinical datasets is required before clinical application.

References

1. American Diabetes Association Professional Practice Committee. (2026a). Diagnosis and classification of diabetes: Standards of Care in Diabetes-2026. Diabetes Care, 49(Suppl. 1), S27-S49. https://doi.org/10.2337/dc26-S002
2. American Diabetes Association Professional Practice Committee. (2026b). Prevention or delay of diabetes and associated comorbidities: Standards of Care in Diabetes-2026. Diabetes Care, 49(Suppl. 1), S50-S60. https://doi.org/10.2337/dc26-S003
3. Baneu, P., Văcărescu, C., Drăgan, S. R., Cirin, L., Lazăr-Höcher, A. I., Cozgarea, A., Faur-Grigori, A. A., Crișan, S., Gaiță, D., Luca, C. T., & Cozma, D. (2024). The Triglyceride/HDL Ratio as a Surrogate Biomarker for Insulin Resistance. Biomedicines, 12(7), 1493. https://doi.org/10.3390/biomedicines12071493
4. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
5. Budak, A., Solmaz, I., Alakuş, Ö. F., & Başgöz, B. B. (2026). Diagnostic Performance of HbA1c for Detecting OGTT-Diagnosed Diabetes in Obese Individuals with Suspected Prediabetes. Journal of clinical medicine, 15(1), 374. https://doi.org/10.3390/jcm15010374
6. Chandrasekaran, P., & Weiskirchen, R. (2024). The Role of Obesity in Type 2 Diabetes Mellitus-An Overview. International journal of molecular sciences, 25(3), 1882. https://doi.org/10.3390/ijms25031882
7. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953
8. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A. L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., Hoffman, M. M., … Logullo, P. (2024). TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ (Clinical research ed.), 385, e078378. https://doi.org/10.1136/bmj-2023-078378
9. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Annals of internal medicine, 162(1), 55–63. https://doi.org/10.7326/M14-0697
10. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273-297. https://doi.org/10.1007/BF00994018
11. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21-27. https://doi.org/10.1109/TIT.1967.1053964
12. Davies, M. J., Lim, S., Slater, T., Goldney, J., Philis-Tsimikas, A., Franco, D. R., Lamptey, R., Yates, T., Tankova, T., & Lingvay, I. (2026). Type 2 diabetes mellitus. Nature

reviews. Disease primers, 12(1), 13. https://doi.org/10.1038/s41572-026-00687-w
13. Silva, K., Lee, W. K., Forbes, A., Demmer, R. T., Barton, C., & Enticott, J. (2020). Use and performance of machine learning models for type 2 diabetes prediction in community settings: A systematic review and meta-analysis. International journal of medical informatics, 143, 104268. https://doi.org/10.1016/j.ijmedinf.2020.104268
14. DeForest, N., Wang, Y., Zhu, Z., Dron, J. S., Koesterer, R., Natarajan, P., Flannick, J., Amariuta, T., Peloso, G. M., & Majithia, A. R. (2024). Genome-wide discovery and integrative genomic characterization of insulin resistance loci using serum triglycerides to HDL-cholesterol ratio as a proxy. Nature communications, 15(1), 8068. https://doi.org/10.1038/s41467-024-52105-y
15. Duncan, B. B., Magliano, D. J., & Boyko, E. J. (2025). IDF Diabetes Atlas 11th edition 2025: global prevalence and projections for 2050. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association, 41(1), 7–9. https://doi.org/10.1093/ndt/gfaf177
16. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010
17. Fregoso-Aparicio, L., Noguez, J., Montesinos, L., & García-García, J. A. (2021). Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. Diabetology & metabolic syndrome, 13(1), 148. https://doi.org/10.1186/s13098-021-00767-9
18. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451
19. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine Learning, 63, 3-42. https://doi.org/10.1007/s10994-006-6226-1
20. International Diabetes Federation. (2025). IDF Diabetes Atlas (11th ed.). International Diabetes Federation.
21. International Expert Committee (2009). International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes care, 32(7), 1327–1334. https://doi.org/10.2337/dc09-9033
22. Kiran, M., Xie, Y., Anjum, N., Ball, G., Pierscionek, B., & Russell, D. (2025). Machine learning and artificial intelligence in type 2 diabetes prediction: a comprehensive 33-year bibliometric and literature analysis. Frontiers in digital health, 7, 1557467. https://doi.org/10.3389/fdgth.2025.1557467
23. Knowler, W. C., Barrett-Connor, E., Fowler, S. E., Hamman, R. F., Lachin, J. M., Walker, E. A., Nathan, D. M., & Diabetes Prevention Program Research Group (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England journal of medicine, 346(6), 393–403. https://doi.org/10.1056/NEJMoa012512
24. Kosmas, C. E., Rodriguez Polanco, S., Bousvarou, M. D., Papakonstantinou, E. J., Peña Genao, E., Guzman, E., & Kostara, C. E. (2023). The Triglyceride/High-Density Lipoprotein Cholesterol (TG/HDL-C) Ratio as a Risk Marker for Metabolic Syndrome and Cardiovascular Disease. Diagnostics (Basel, Switzerland), 13(5), 929. https://doi.org/10.3390/diagnostics13050929
25. Lee, H., Hwang, S. H., Park, S., Choi, Y., Lee, S., Park, J., Son, Y., Kim, H. J., Kim, S., Oh, J., Smith, L., Pizzol, D., Rhee, S. Y., Sang, H., Lee, J., & Yon, D. K. (2025). 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. https://doi.org/10.1016/j.eclinm.2025.103069
26. Lugner, M., Rawshani, A., Helleryd, E., & Eliasson, B. (2024). Identifying top ten predictors of type 2 diabetes through machine learning analysis of UK Biobank data. Scientific reports, 14(1), 2102. https://doi.org/10.1038/s41598-024-52023-5
27. Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., Oakden-Rayner, L., … van Smeden, M. (2025). PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ (Clinical research ed.), 388, e082505. https://doi.org/10.1136/bmj-2024-082505
28. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
29. Putri, R. R., Danielsson, P., Hagman, E., & Marcus, C. (2026). Long-Term Cardiometabolic Outcomes in Children with Metabolically Healthy and Unhealthy Obesity. JAMA pediatrics, e260343. Advance online publication. https://doi.org/10.1001/jamapediatrics.2026.0343
30. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533-536. https://doi.org/10.1038/323533a0
31. Sacks, D. B., Arnold, M., Bakris, G. L., Bruns, D. E., Horvath, A. R., Lernmark, Å., Metzger, B. E., Nathan, D. M., & Kirkman, M. S. (2023). Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus. Diabetes care, 46(10), e151–e199. https://doi.org/10.2337/dci23-0036
32. Sulu, C., & Yumuk, V. D. (2024). Treat Obesity to Treat Type 2 Diabetes Mellitus. Diabetes therapy : research, treatment and education of diabetes and related disorders, 15(3), 611–622. https://doi.org/10.1007/s13300-024-01536-3
33. Tasnia, R., & Obeng-Gyasi, E. (2026). Evaluating Machine Learning Models for Classifying Diabetes Using Demographic, Clinical, Lifestyle, Anthropometric, and

Environmental Exposure Factors. Toxics, 14(1), 76. https://doi.org/10.3390/toxics14010076

34. Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., Mallett, S., & PROBAST Group† (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Annals of internal medicine, 170(1), 51–58. https://doi.org/10.7326/M18-1376