Received: 29 September 2022 | Revised: 15 November 2022 | Accepted: 21 November 2022

DOI: 10.1002/jcla.24798

RESEARCH ARTICLE

WILEY

Data mining approaches for type 2 diabetes mellitus prediction using anthropometric measurements

Maryam Saberi-Karimian[1] | Amin Mansoori[1,2] | Maryam Mohammadi Bajgiran[1] | Zeinab Sadat Hosseini[3] | Amir Kiyoumarsioskouei[4] | Elias Sadooghi Rad[5,6] | Mostafa Mahmoudi Zo[5] | Negar Yeganeh Khorasani[5] | Mohadeseh Poudineh[5,7] | Sara Ghazizadeh[8] | Gordon Ferns[9] | Habibollah Esmaily[2,10] | Majid Ghayour-Mobarhan[1]

[1]International UNESCO center for Health Related Basic Sciences and Human Nutrition, Mashhad University of Medical Sciences, Mashhad, Iran

[2]Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran

[3]Faculty of Medicine, Islamic Azad University of Mashhad, Mashhad, Iran

[4]Faculty of Mechanical Engineering, Sahand University of Technology, Tabriz, Iran

[5]Student Research Committee, School of Medicine, Mashhad University of Medical sciences, Mashhad, Iran

[6]Student Research Committee, School of Medicine, Birjand University of Medical sciences, Birjand, Iran

[7]School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran

[8]Department of Biology, Faculty of Sciences, Mashhad Branch, Islamic Azad University, Mashhad, Iran

[9]Brighton and Sussex Medical School, Division of Medical Education, Brighton, UK

[10]Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

Correspondence
Majid Ghayour-Mobarhan, International UNESCO center for Health Related Basic Sciences and Human Nutrition, Department of Nutrition, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
Email: ghayourm@mums.ac.ir

Habibollah Esmaily, Social Determinants of Health Research Center, Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.
Email: esmailyh@mums.ac.ir

Abstract

Background: The aim of this study was to evaluate the anthropometric measurements most associated with type 2 diabetes mellitus (T2DM) using machine learning approaches.

Methods: A prospective study was designed for a total population of 9354 (43% men and 57% women) aged 35–65. Anthropometric measurements include weight, height, demispan, Hip Circumference (HC), Mid-arm Circumference (MAC), Waist Circumference (WC), Body Roundness Index (BRI), Body Adiposity Index (BAI), A Body Shape Index (ABSI), Body Mass Index (BMI), Waist-to-height Ratio (WHtR), and Waist-to-hip Ratio (WHR) were completed for all participants. The association was assessed using logistic regression (LR) and decision tree (DT) analysis. Receiver operating characteristic (ROC) curve was performed to evaluate the DT's accuracy, sensitivity, and specificity using R software.

Results: Traditionally, 1461 women and 875 men with T2DM (T2DM group). According to the LR, in males, WC and BIA (p-value <0.001) and in females, demispan and WC

Maryam Saberi-Karimian, Amin Mansoori, and Maryam Mohammadi Bajgiran are equally first author.

(p-value < 0.001) had the highest correlation with T2DM development risk. The DT indicated that WC has the most crucial effect on T2DM development risk, followed by HC, and BAI.

Conclusions: Our results showed that in both men and women, WC was the most important anthropometric factor to predict T2DM.

KEYWORDS
anthropometric, data mining, decision tree, diabetes

1 | INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a chronic disease that is diagnosed by an abnormal blood glucose level due to insulin deficiency or insulin resistance.[1] However, recent studies suggest that the accumulation of ectopic lipids may play an important role in developing T2DM.[2] The International Diabetes Federation reports that the global prevalence of diabetes in adults in 2021 was 536.6 million people, and it is estimated that it will rise to 783.2 million in 2045.[3] In Iran, a 14.3% prevalence of T2DM was reported in 2019,[4] this has emerged as a great concern, not only because of the numbers but also because of microvascular or macrovascular complications caused by T2DM.[5] It should be emphasized that one out of nine deaths occur due to diabetes.[6] Despite all efforts to treat T2DM and its complications,[7,8] prevention strategies are still the best way to control the disease. There is a growing body of literature suggesting that preventing obesity in the population can decrease the risk of diabetes.[9] Recent evidence suggests that patients with metabolic syndrome more frequently develop T2DM, which is principally linked to abdominal or visceral obesity and metabolic abnormalities.[2] Considering both the prevalence and the crucial complications associated with T2DM, it is very important to identify people who are at great risk for T2DM. While one-third of patients with diabetes have been undiagnosed and this event is increasing,[10] anthropometric measurement techniques are one of the simplest and non-invasive methods of identifying people at risk of diabetes.[11]

Waist-to-height ratio (WHtR), Body Adiposity Index (BAI), A Body Shape Index (ABSI), Body Mass Index (BMI), and Waist Circumference (WC) are predictors contributing to diabetes.[12,13] One study reported that WC and BMI are the best predictors,[12] but another study suggested WC and WHtR are better than BMI.[14] Furthermore, a study indicated that WHtR is the strongest predictor,[13] but there are few studies evaluating all of them. And also, there were few studies about the relation of other anthropometric measurements such as demispan and Mid-Arm Circumference (MAC) with T2DM.[15] Despite the many studies that have been done on this subject, few of them have compared the values of these anthropometric indices and their relationship with the prevalence of diabetes. Also, traditional anthropometric methods do not detect visceral and subcutaneous fat. As these factors have different roles in predicting diabetes and few studies have investigated the sex-specific difference in the impact of anthropometric changes on the risk of diabetes in general populations, we need to find a strong way to use them for many groups of people.[1,14]

According to the increasing prevalence of diabetes and the lack of reliable measurements for diagnosing the disease, and also the controversy of previous study results, we decided to design a cohort study to examine more and new anthropometrical measurements for evaluating their association with diabetes. We also used machine learning approaches which are new and effective methods for arranging a large number of predictors while producing strong predictive models to identify subjects at risk of developing diseases.

2 | METHOD

2.1 | Study population

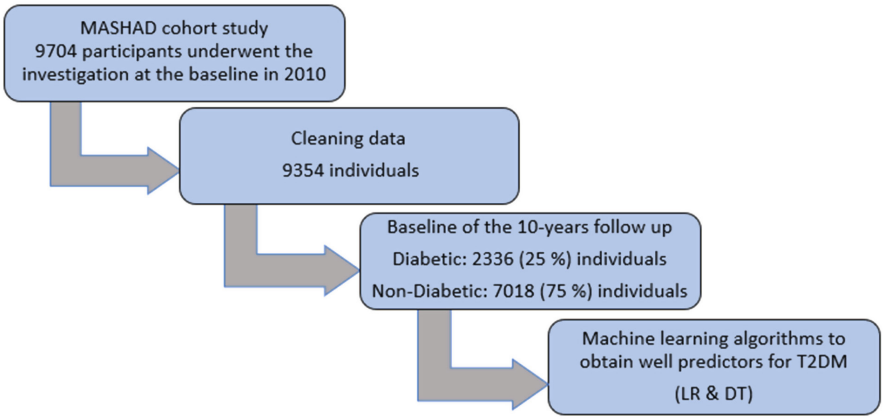
All participants were included from the baseline of the Mashhad stroke and heart atherosclerotic disorder (MASHAD) study, a 10-year cohort from northeastern Iran.[16] The prevalence of diabetes was estimated to be 3% in the Mashhad population based on the data from the Ministry of Health. Subjects were registered from three districts of Mashhad. Each district was divided into nine areas centered at Mashhad Healthcare Center divisions. The baseline investigation was started in 2010 obtaining a response of 79% after stratified cluster random sampling. After recognizing eligible subjects, we arranged a meeting for the physical examination. The inclusion criteria were males and females between the age of 35 and 65 years. Also, those who were not between the ages of 35 and 65 and did not consent to participate in the study were excluded from the present analysis.

Accordingly, 9704 individuals aged 35–65 years were enrolled. After cleaning the data, we had 9354 individual datasets (see Figure 1). All participants completed written consent forms, and the study protocol was approved by the Ethical Committee of Mashhad University of Medical Sciences.

2.2 | Baseline examination

Subjects were asked for taking blood samples between 8 and 10 a.m. by venipuncture of an antecubital vein after 14 h of

FIGURE 1 Flow chart of this study



overnight fasting. According to a standard protocol, the specimens were collected in vacuum tubes (20 ml) in a sitting position from participants. All blood samples were centrifuged at 25 °C within 30–45 min of collection to divide the serum from plasma into six aliquots (0.5 ml). Then they were transported to the Bu Ali Research Institute, Mashhad. Also, for future analysis, aliquots of serum were kept frozen at −80 °C. T2DM was determined as fasting blood glucose (FBG) ≥ 126 mg/dl or being treated with available oral hypoglycemic medications or insulin. Anthropometric measurements such as weight, height, demispan, HC (Hip Circumference), MAC (Mid-arm Circumference), WC (Waist Circumference), BRI (Body Roundness Index), BAI (Body Adiposity Index), ABSI (A Body Shape Index), BMI (Body Mass Index), WHtR (waist-to-height Ratio), and WHR (Waist-to-hip Ratio) were measured by a registered nurse. The subjects were asked to wear light clothes and no shoes during measurements of height and weight. Height and weight were measured with a fixed stadiometer calibrated in centimeters to the closest 0.1 cm and electronic scales to the nearest 0.1 kg, respectively. Demispan was defined as the distance between the mid-sternal notch and the space between the middle and ring fingers in a stretched arm. HC was calculated at the biggest circumference between the crotch and the iliac crest. WC was measured at the middle point between the iliac crest and the last rib. Subjects assessed in a standing position and were asked to exhale during measurements of HC and WC by using a flexible and inelastic calibrated tape measure. According to International Diabetes Federation, WC > 94 cm and >80 cm were considered high as central obesity for males and females, respectively. MAC was described as the distance between the point of the elbow and the bony protrusion on the shoulder while the subject's elbow was bent 90 degrees. BRI, BAI, and ABSI were calculated using the formulas in Table 1 in Appendix S1. BMI was defined as the weight (kg) divided by the square of height (m). Based on the World Health Organization recommendations, a BMI ≥ 25–29.99 was defined as overweight and a BMI ≥30 kg/m^2 was considered obese.[17] WHR was calculated as WC divided by HC. As reported by the World Health Organization recommendations, truncal obesity was defined as a *WHR* > 0.95 in males and a *WHR* > 0.8 in females.[18] WHtR was measured by dividing WC (m) based on height (m).[19]

2.3 | Statistical analysis

To describe the quantitative and qualitative variables, mean ± SD and frequency (%) were reported, respectively. Chi-square and Fisher's exact tests were applied to measure the association between qualitative variables. Also, the mean of quantitative variables between the two groups were compared by independent *t*-test. All of the analyses were done separately for males and females. Multiple logistic regression is applied to investigate the relation between anthropometric measurements and T2DM. Also, at first univariate LR was fitted to the data and then variables with *p*-value < 0.05 were entered in multiple LR model. Also, their odds ratios (OR) were calculated. The version of the SPSS program was 22 (SPSS Inc.). *p*-Value < 0.05 was regarded as significant.

2.4 | Decision tree model

We put the data into machine learning, and the decision tree was drawn to form a predictive model of anthropometric measurements. A decision tree is a non-parametric method named regarding the nature of the target variable. The aim of a decision tree is to form a predictive model in terms of predictor variables. In order to form the tree, decision tree algorithms make splitting criteria at internal nodes.[20,21] The split of a node tries to minimize the impurity of the node. The node is not split and is characterized as a leaf node if a split is not able to achieve any enhancements with regard to reducing impurity. If a split manages to decrease impurity, then the split having the maximum reduction in impurity is chosen and two branches are formed, making two new nodes. The most popular splitting criteria are the Gini index:

$$Gini(D) = 1 - \sum_{i=1}^{m} P_i^2$$

CART is a decision tree algorithm that builds a binary tree using the Gini index for choosing the splitting variable at each internal node. The tree starts with all observations forming the root node and successive splits specify the order of significance of the predictor variables. Receiver operating characteristic (ROC) curves were used to evaluate the accuracy, precision, and specificity of the decision tree algorithm using R software version 4.0.5. And then the confusion matrix of the decision trees was evaluated.

Two data mining techniques were used to investigate the relationship between anthropometric predictors and binary response variables (diabetic and non-diabetic). So, the main objective of this study was to anticipate diabetes using the LR and DT models and to determine their associated factors, especially anthropometric markers. For this purpose, the dataset was randomly split into two parts: training data and test data (20%–80%). The training dataset was utilized to develop the LR and DT models, which was then validated using test data (20%) that had not been used during training.

3 | RESULT

3.1 | Characteristics of the study population

The study population consisted of 57% females and 43% male individuals ($N = 9354$ in total). The study population sample included 1461 women individuals with T2DM (T2DM group) and 3938 samples without T2DM. In men, 3080 subjects did not have T2DM, and 875 subjects had T2DM throughout the study. The clinical characteristics of the participants at the baseline have been summarized in Table 1.

3.2 | The association between anthropometric measurements and T2MD using logistic regression (LR) model

According to the data mining analysis results (Table 2 LogWorth and *p*-Value columns), BMI, WC, HC, and BAI had a significant association with T2DM in males and demispan, HC, WC, BMI, and BAI had a significant association with T2DM in females (p-value < 0.05). In males, WC and BAI had the greatest correlation with the development of T2DM among the analyzed anthropometric factors. Also, in

TABLE 1 Baseline characteristics of male and female

Male			
Variables	**Diabetes+ (875)**	**Diabetes− (3080)**	***p*-Value**
BMI (kg/m^2)	27.52 ± 4.25	26.90 ± 4.26	<0.001
WC (cm)	97.49 ± 11.35	95.10 ± 10.87	<0.001
HC (cm)	102.34 ± 9.03	101.46 ± 8.69	0.009
WHR	0.95 ± 0.07	0.93 ± 0.06	<0.001
MAC (cm)	30.54 ± 3.51	29.81 ± 4.11	<0.001
BAI	29.66 ± 4.22	28.68 ± 4.65	<0.001
WHtR	0.58 ± 0.06	0.56 ± 0.06	<0.001
ABSI	0.083 ± 0.004	0.082 ± 0.005	<0.001
BRI	5.19 ± 1.40	4.78 ± 1.45	<0.001
Demispan (cm)	79.73 ± 4.64	80.76 ± 5.06	<0.001
Female			
Variables	**Diabetes + (1461)**	**Diabetes− (3938)**	***p*-Value**
BMI (kg/m^2)	29.66 ± 4.50	29.19 ± 4.85	0.001
WC (cm)	101.49 ± 11.58	96.97 ± 11.99	<0.001
HC (cm)	105.03 ± 9.19	105.93 ± 9.60	0.002
WHR	0.96 ± 0.07	0.91 ± 0.08	<0.001
MAC (cm)	30.99 ± 3.62	30.59 ± 3.59	<0.001
BAI	36.90 ± 5.49	36.95 ± 5.65	0.021
WHtR	0.65 ± 0.07	0.62 ± 0.08	<0.001
ABSI	0.08 ± 0.007	0.08 ± 0.008	<0.001
BRI	6.92 ± 1.89	6.15 ± 1.90	<0.001
Demispan (cm)	74.33 ± 7.10	73.92 ± 4.23	<0.001

Abbreviations: ABSI, A Body Shape Index; BMI, Body Mass Index; BAI, Body Adiposity Index; BRI, Body Roundness Index; HC, Hip Circumference; WC, Waist Circumference; WHR, Waist-to-hip Ratio; WHtR, Waist-to-height Ratio.

females, demispan and WC had the highest correlation with T2DM development among the analyzed anthropometric factors.

Table 2 (OR column) shows the unit odds ratios based on the significant factors. In males, WC and BAI were significantly associated with T2DM. Among these factors, BAI has been identified as the most remarkable risk factor for T2DM (OR: 1.078, (95% CI: 1.044, 1.113)). In females, demispan, WC, and BAI were significantly associated with T2DM. Among these factors, WC has been identified as the most remarkable risk factor for T2DM (OR: 1.085, (95% CI: (1.074, 1.095))).

TABLE 2 Parameter estimates of the LR model for T2DM by anthropometric factors in male and female

Male			
Terms	***p*-Value**	**LogWorth**	**OR (CI 95%)**
BMI (kg/m^2)	0.0156	1.806	0.951 (0.912, 0.991)
HC (cm)	<0.0001	4.508	0.958 (0.939, 0.977)
WC (cm)	<0.0001	6585	1.042 (1.025, 1.059)
BAI	<0.0001	5.564	1.078 (1.044, 1.113)
Female			
Terms	***p*-Value**	**LogWorth**	**OR (95% CI)**
Demispan (cm)	<0.0001	9.281	1.053 (1.036, 1.069)
WC (cm)	<0.0001	68.330	1.085 (1.074, 1.095)
HC (cm)	<0.0001	45.266	0.887 (0.872, 0.902)
BMI (kg/m^2)	0.0256	1.591	1.035 (1.004, 1.066)
BAI	0.0001	3.728	1.051 (1.024, 1.078)

Abbreviations: BAI, Body Adiposity Index; BMI, Body Mass Index; HC, Hip Circumference; WC, Waist Circumference.

3.3 | The association between anthropometric measurements and T2DM using decision tree (DT) model

The results of the DT training for anthropometric factors in males is shown in Figure 2. The DT algorithm evaluated the various T2DM risk factors and categorized them into three layers. In the DT model, the first variable (root) is of the highest importance, with the following variables in the next levels of significance, accordingly. As shown in Figure 2, WC has the most crucial effect on T2DM development risk, followed by BAI, and HC. In the subgroup with WC ≥ 94, HC < 95.5, and WC ≥ 97, 93% of participants were diabetic (highest risk of T2DM). Meanwhile, among those with WC < 94, BAI < 30.46, and HC ≥ 96.7, 95% of subjects were identified as non-diabetic (lowest risk of T2DM). Detailed rules for T2DM created by the DT model are demonstrated in Table 3.

The results of the DT training for anthropometric factors in females are demonstrated in Figure 3. The DT algorithm evaluated the various T2DM risk factors and categorized them into four layers. In the DT model, the first variable (root) is of the highest importance, with the following variables in the next levels of significance, accordingly. As shown in Figure 3, WC has the most crucial effect on T2DM development risk, followed by HC, demispan, BAI, and BMI. In the subgroup with WC ≥ 106, HC < 109.5, and Demispan ≥ 81.2, 98% of participants were diabetic (highest risk of T2DM). Meanwhile, among those with WC < 106, HC ≥ 102, BAI < 39.69, and WC < 102, 95% of subjects were identified as non-diabetic (lowest risk of T2DM). Detailed rules for T2DM created by the DT model are demonstrated in Table 3.

In order to evaluate the performance of the model and comparisons, we gave the receiver operating characteristics (ROC) curve and the confusion matrix of the algorithm for both training and testing data (Figure 4 and Table 4, respectively). Table 4 shows how

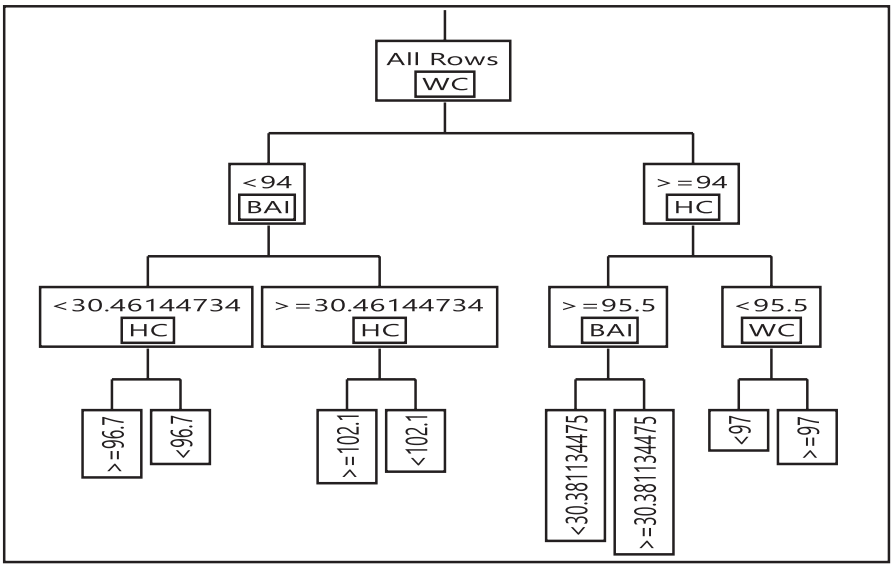


FIGURE 2 Decision tree for type 2 diabetes mellitus event in male

remarkable the model performs, since all measures in both tables (test vs. training) are almost identical which is also confirmed by Figure 4.

4 | DISCUSSION

Obesity is a condition of increasing prevalence across the world, with important social and economical implications, being recognized as a public health concern.[22] Evidence shows that obesity is a major risk factor for T2DM, for instance nearly 85% of the US population with T2DM are either overweight or obese.[23] Anthropometric measurements are useful and cheap indicators to determine the risk of T2DM. Our comprehensive study has attempted to define the best anthropometric parameters for the prediction of T2DM in the northeastern Iranian population.

According to our study, the Incidence Rate (IR) of T2DM was higher in females (IR = 27%) than in males (IR = 22%). It may be due to the fact that women in Iran generally participate in less physical activities and most of the time they are at home. Both in males and females, the most associated factors with T2DM were BMI, HC, BAI, and WC. Moreover, in men BAI and WC and in women demispan, WC, BMI, and BAI increased the risk of T2DM. Among these factors, WC in both males and females was the most remarkable factors. On the other hand, there was not a strong relationship between HC and BMI in males and HC in females with increasing T2DM. Furthermore, by using machine learning algorithms, we concluded that the best factors to categorize data are BAI, HC and WC in males and females, respectively.

Obesity is a major risk factor for T2DM. But studies suggest that BMI is more of a precipitating factor than a trigger factor for the development of T2DM.[24] The literature suggests that BMI is a good predictor in screening for T2DM.[12] In contrast, the results of some studies showed that predicted fat mass and Visceral Adiposity Index (VAI) are more significantly associated with the risk of T2DM than BMI.[25,26] In our study, BMI is a remarkable risk factor for T2DM in females. Surprisingly, BMI was found to be inversely associated with T2DM in men. This observation may be explained by the fact that men have more insulin resistant than women because of their more visceral and liver adipose tissue, as well as the absence of estrogen protective effect.[27,28] Another explanation could be that BMI overestimates body fat mass in men; because women have a greater amount of total body fat.[29] Also, it has been found that middle-aged men usually have higher triglyceride and fasting blood glucose levels

TABLE 3 Detailed rules based on DT model for male and female

Male			
Num	**Rules**	**Diabetic (%)**	**Non-diabetic (%)**
1	WC < 94 & BAI < 30.46 & HC > =96.7	4.92	95.08
2	WC < 94 & BAI < 30.46 & HC < 96.7	19.17	80.83
3	WC < 94 & BAI > =30.46 & HC > =102.1	04.42	95.58
4	WC < 94 & BAI > =30.46 & HC < 102.1	56.29	43.71
5	WC > =94 & HC > =95.5 & BAI < 30.38	13.83	86.17
6	WC > =94 & HC > =95.5 & BAI > =30.38	33.06	66.94
7	WC > =94 & HC < 95.5 & WC < 97	6.86	93.14
8	WC > =94 & HC < 95.5 & WC > =97	92.71	7.29
Female			
Num	**Rules**	**Diabetic (%)**	**Non-diabetic (%)**
1	WC < 106 & HC > =102 & BAI < 39.69 & WC < 102	5.51	94.49
2	WC < 106 & HC > =102 & BAI < 39.69 & WC > =102	29.87	70.13
3	WC < 106 & HC > =102 & BAI > =39.69	20.48	79.52
4	WC < 106 & HC < 102 & BMI < 22.18	4.64	95.36
5	WC < 106 & HC < 102 & BMI > =22.18 & Demispan<89.5	33.41	66.59
6	WC < 106 & HC < 102 & BMI > =22.18 & Demispan> = 89.5	93.36	6.64
7	WC > =106 & HC > =109.5 & BMI < 30.88	5.56	94.44
8	WC > =106 & HC > =109.5 & BMI > =30.88 & BMI > =31.78	27.23	72.77
9	WC > =106 & HC > =109.5 & BMI > =30.88 & BMI < 31.78	80.27	19.73
10	WC > =106 & HC < 109.5 & Demispan<81.2 & Demispan> = 73.5	19.94	80.06
11	WC > =106 & HC < 109.5 & Demispan<81.2 & Demispan<73.5	67.26	32.74
12	WC > =106 & HC < 109.5 & Demispan> = 81.2	98.59	1.41

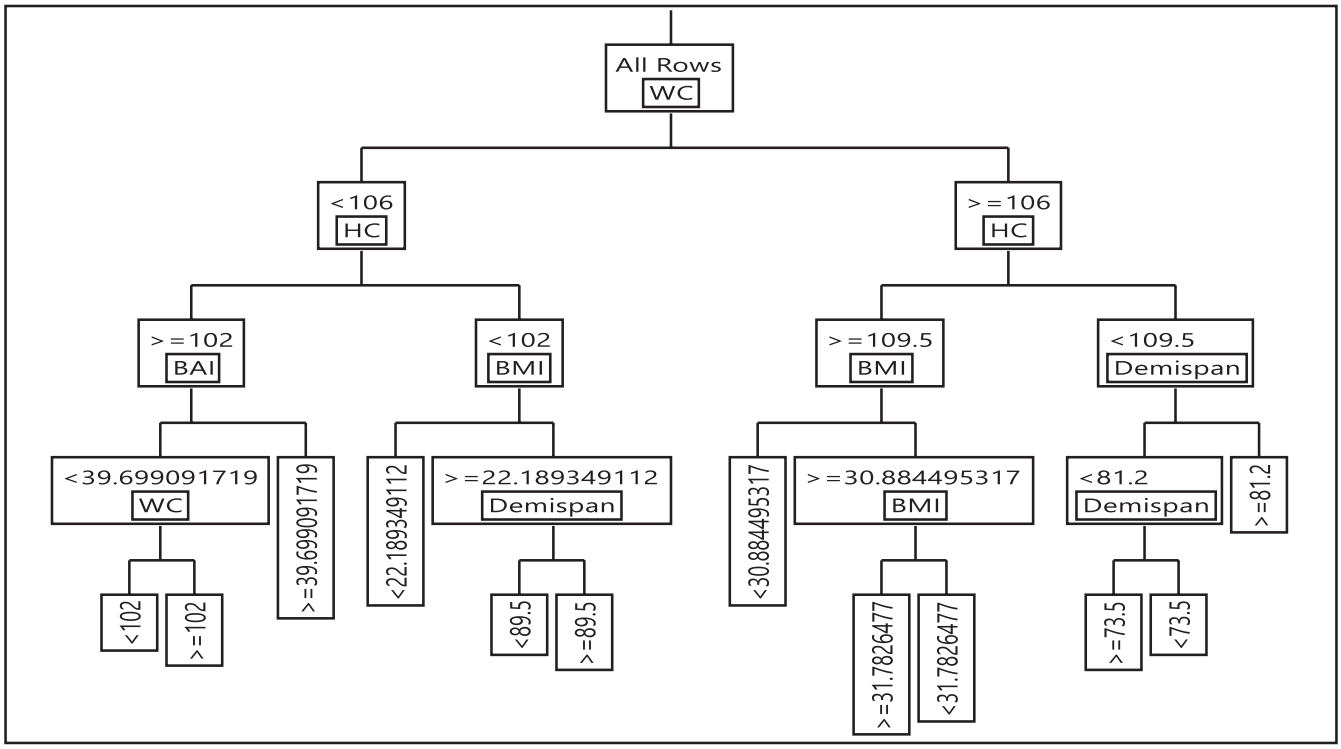


FIGURE 3 Decision tree for type 2 diabetes mellitus event in female

and lower HDL cholesterol levels than women of the same age even after adjusting for BMI.[30,31] Although more studies should be done to understand the precise mechanisms, there may be other factors such as different lifestyles or other hormonal differences between men and women.[28] According to a meta-analysis study, there is a strong inverse association between HC and T2DM in both men and women.[32] The possible explanation for these results regarding BMI and HC in our study is explained by the fact that the distribution of adipose tissue is a more important predicting factor than the absolute amount of fat in the body. Some studies stated that high WC can increase the risk of T2DM.[14,33] We agree in terms of WC is a risk factor for diabetes. To the best of our knowledge, this was the first study to evaluate the relationship between demispan and T2DM. This study confirms that demispan is associated with diabetes in women, increasing the risk. In accordance with the present results, previous studies have demonstrated that higher BAI increases the risk of T2DM.[34–36]

4.1 | Strengths and limitations

Cohort studies may better identify factors that are causally related to a particular disease. Our study suggests a causal relationship between anthropometric measurements and T2DM. Furthermore, its large population-based sample of northeastern Iran was another strength. In this study, we used new analyzing methods including machine learning algorithms such as decision tree to arrange T2DM predictors. For interpreting our results there are the following limitations: Nearly half of all individuals with T2DM were older adults (aged ≥ 65 years) but we only included subjects aged between 35 and 65, so we might have lost many potential cases of T2DM that could affect our data analysis.[37] Another one is that we only used the FBG test to diagnose T2DM, so we might have made either overdiagnosis or under diagnosis of T2DM. We recommend for further works using more confirmatory tests such as HbA1c and 2 h-Glucose Tolerance. We only evaluated the relation between anthropometric measurements with T2DM but other factors such as epigenetics, family history of T2DM, lipid profile, inflammatory markers, and other confounding factors that can interact with anthropometric profile were not examined. Future studies on the current topic are therefore recommended to examine the relationship between these factors and anthropometric measurements with T2DM at the same time. The other limitation was the imbalanced data that we faced in this study. In this case, some evaluation metrics such as sensitivity are affected. For future works suggest analyzing the balanced data.

5 | CONCLUSION

Our results showed that in both men and women WC was the strongest predictors of T2DM. Also, both in men and women BAI, and WC were the most associated factors with T2DM. DT analysis identified BAI, HC, and WC as the best variables to categorize

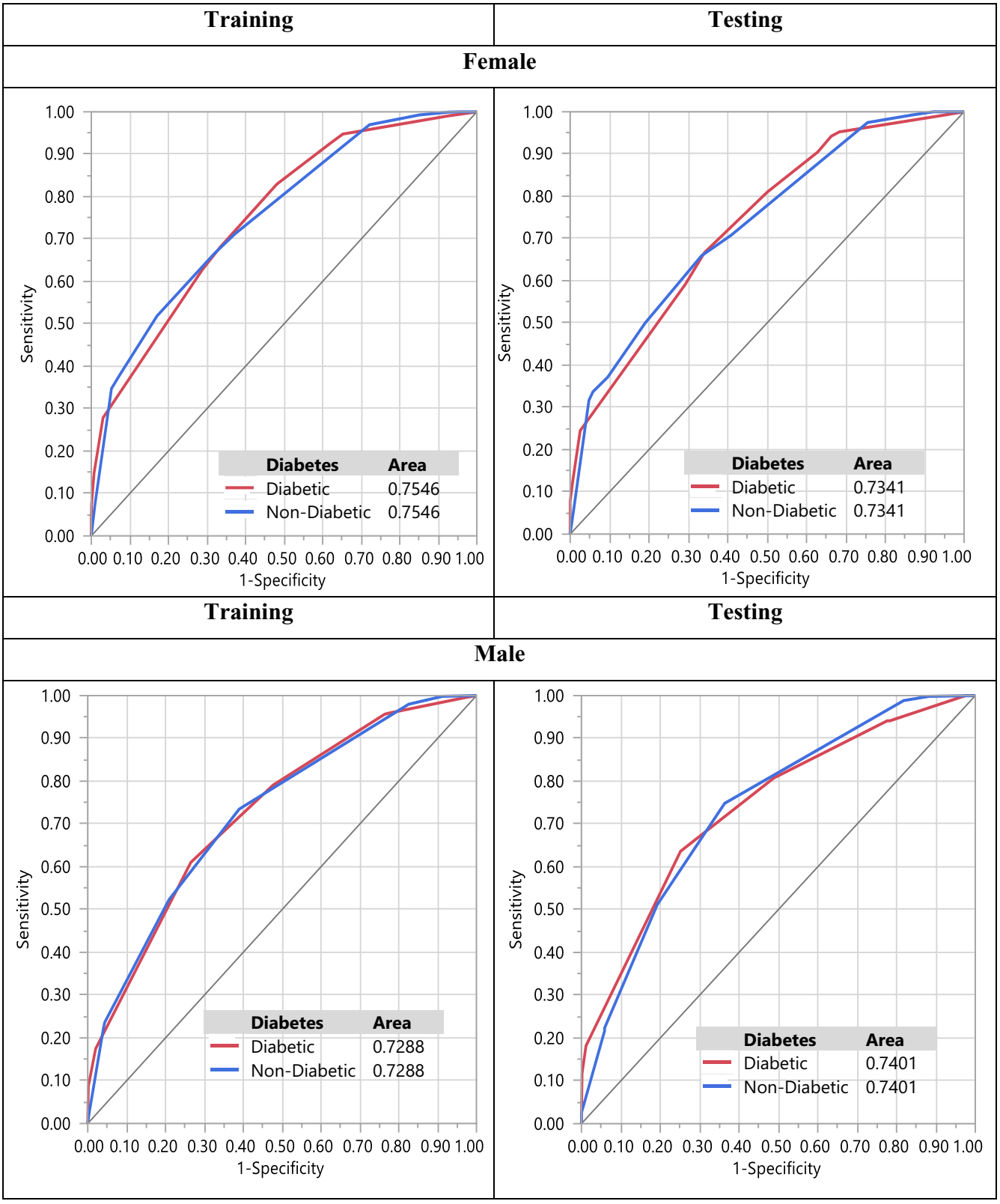


FIGURE 4 ROC curve of DT model for male and female

TABLE 4 Performance indices of the LR model for male and female

Female

(a) Training (n= 4319)

Actual	Predicted Count	
	Diabetic	Non- Diabetic
Diabetic	325	842
Non- Diabetic	96	3056

Specificity = 96.95% | AUC = 75.46%

Accuracy = 78.28% | Precision = 77.19%

(b) Testing (n= 1080)

Actual	Predicted Count	
	Diabetic	Non- Diabetic
Diabetic	72	222
Non- Diabetic	20	766

Specificity = 97.45% | AUC = 73.41%

Accuracy = 77.59% | Precision = 78.26%

Male

(c) Training (n= 3164)

Actual	Predicted Count	
	Diabetic	Non- Diabetic
Diabetic	120	568
Non- Diabetic	50	2426

Specificity = 97.98% | AUC = 72.88%

Accuracy = 80.46% | Precision = 70.58%

(d) Testing (n= 791)

Actual	Predicted Count	
	Diabetic	Non- Diabetic
Diabetic	34	153
Non- Diabetic	7	597

Specificity = 98.84% | AUC = 74.01%

Accuracy = 79.77% | Precision = 82.92%

data in men and women. Therefore, it was recommended to use these anthropometric factors for predicting the risk and screening of T2DM.

AUTHOR CONTRIBUTIONS

Habibollah Esmaily and Majid Ghayour Mobarhan are corresponding author. Maryam Saberi-Karimiam conceptualized and revised the article. Amin Mansoori conceptualized the study and analyzed the data. Maryam Mohammadi Bajgiran contributed to formal analysis. Zeinab Sadat Hosseini conceptualized and drafted the article. Amir Kiyoumarsioskouei analyzed the data. Elias Sadooghi Rad, Mostafa Mahmoudi Zo, Negar Yeganeh Khorasani, and Sara Ghazizadeh drafted the article. Mohadeseh Poudineh drafted and revised the article. Gordon Ferns revised the article.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ETHICAL APPROVAL

This study protocol was reviewed and approved by the Ethics Committee of MUMS, approval number IR.MUMS.REC.1386.250.

ORCID

Majid Ghayour-Mobarhan https://orcid.org/0000-0002-1081-6754

REFERENCES

1. Amiri P, Javid AZ, Moradi L, et al. Associations between new and old anthropometric indices with type 2 diabetes mellitus and risk of metabolic complications: a cross-sectional analytical study. *J Vasc Bras*. 2021;20:e20200236.
2. Bozorgmanesh M, Hadaegh F, Azizi F. Diabetes prediction, lipid accumulation product, and adiposity measures; 6-year follow-up: Tehran lipid and glucose study. *Lipids Health Dis*. 2010;9:45.
3. Sun H, Saeedi P, Karuranga S, et al. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract*. 2022;183:109119.
4. Ranjbaran S, Shojaeizadeh D, Dehdari T, Yaseri M, Shakibazadeh E. The effectiveness of an intervention designed based on health action process approach on diet and medication adherence among patients with type 2 diabetes: a randomized controlled trial. *Diabetol Metab Syndr*. 2022;14(1):3.
5. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol*. 2020;16(7):377-390.
6. Saeedi P, Salpea P, Karuranga S, et al. Mortality attributable to diabetes in 20-79 years old adults, 2019 estimates: results from the international diabetes federation diabetes atlas, 9(th) edition. *Diabetes Res Clin Pract*. 2020;162:108086.
7. Sattar N, Lee MMY, Kristensen SL, et al. Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of randomised trials. *Lancet Diabetes Endocrinol*. 2021;9(10):653-662.
8. Akbari A, Rafiee M, Sathyapalan T, Sahebkar A. Impacts of sodium/glucose Cotransporter-2 inhibitors on circulating uric acid concentrations: a systematic review and meta-analysis. *J Diabetes Res*. 2022;2022:7520632-7520617.

9. Borné Y, Nilsson PM, Melander O, Hedblad B, Engström G. Multiple anthropometric measures in relation to incidence of diabetes: a Swedish population-based cohort study. *Eur J Public Health*. 2015;25(6):1100-1105.
10. Mirahmadizadeh A, Fathalipour M, Mokhtari AM, Zeighami S, Hassanipour S, Heiran A. The prevalence of undiagnosed type 2 diabetes and prediabetes in eastern Mediterranean region (EMRO): a systematic review and meta-analysis. *Diabetes Res Clin Pract*. 2020;160:107931.
11. Zhao Q, Zhang K, Li Y, et al. Capacity of a body shape index and body roundness index to identify diabetes mellitus in Han Chinese people in Northeast China: a cross-sectional study. *Diabet Med*. 2018;35(11):1580-1587.
12. Wei J, Liu X, Xue H, Wang Y, Shi Z. Comparisons of visceral adiposity index, body shape index, body mass index and waist circumference and their associations with diabetes mellitus in adults. *Nutrients*. 2019;11(7):1580.
13. Zhang F, Wan Q, Cao H, et al. Identical anthropometric characteristics of impaired fasting glucose combined with impaired glucose tolerance and newly diagnosed type 2 diabetes: anthropometric indicators to predict hyperglycaemia in a community-based prospective cohort study in Southwest China. *BMJ Open*. 2018;8(5):e019735.
14. Zhang FL, Ren JX, Zhang P, et al. Strong Association of Waist Circumference (WC), body mass index (BMI), waist-to-height ratio (WHtR), and waist-to-hip ratio (WHR) with diabetes: a population-based cross-sectional study in Jilin Province, China. *J Diabetes Res*. 2021;2021:8812431.
15. Zhu Y, Lin Q, Zhang Y, et al. Mid-upper arm circumference as a simple tool for identifying central obesity and insulin resistance in type 2 diabetes. *PLoS One*. 2020;15(5):e0231308.
16. Ghayour-Mobarhan M, Moohebati M, Esmaily H, et al. Mashhad stroke and heart atherosclerotic disorder (MASHAD) study: design, baseline characteristics and 10-year cardiovascular risk estimation. *Int J Public Health*. 2015;60(5):561-572.
17. 2. Classification and diagnosis of diabetes: standards of medical Care in Diabetes-2018. *Diabetes Care*. 2018;41(Suppl 1):S13-s27.
18. World Health Organization. *Regional Office for the Western P. the Asia-Pacific Perspective: Redefining Obesity and its Treatment*. Vol 2000. Health Communications Australia; 2000.
19. Ashwell M, Gibson S. Waist to height ratio is a simple and effective obesity screening tool for cardiovascular risk factors: analysis of data from the British National Diet and Nutrition Survey of adults aged 19-64 years. *Obes Facts*. 2009;2(2):97-103.
20. Aghasizadeh M, Samadi S, Sahebkar A, et al. Serum HDL cholesterol uptake capacity in subjects from the MASHAD cohort study: its value in determining the risk of cardiovascular endpoints. *J Clin Lab Anal*. 2021;35(6):e23770.
21. Saberi-Karimian M, Safarian-Bana H, Mohammadzadeh E, et al. A pilot study of the effects of crocin on high-density lipoprotein cholesterol uptake capacity in patients with metabolic syndrome: a randomized clinical trial. *Biofactors*. 2021;47(6):1032-1041.
22. Ortega MA, Fraile-Martínez O, Naya I, et al. Type 2 diabetes mellitus associated with obesity (Diabesity). The central role of gut microbiota and its translational applications. *Nutrients*. 2020;12(9):2749.
23. Adult Obesity Facts 2015. https://www.cdc.gov/obesity/data/adult.html
24. Vázquez JA, Gaztambide S, Soto-Pedre E. 10-year prospective study on the incidence and risk factors for type 2 diabetes mellitus. *Med Clin (Barc)*. 2000;115(14):534-539.
25. Lee DH, Keum N, Hu FB, et al. Comparison of the association of predicted fat mass, body mass index, and other obesity indicators with type 2 diabetes risk: two large prospective studies in US men and women. *Eur J Epidemiol*. 2018;33(11):1113-1123.
26. Koloverou E, Panagiotakos DB, Kyrou I, et al. Visceral adiposity index outperforms common anthropometric indices in predicting 10-year diabetes risk: results from the ATTICA study. *Diabetes Metab Res Rev*. 2019;35(6):e3161.
27. Geer EB, Shen W. Gender differences in insulin resistance, body composition, and energy balance. *Gend Med*. 2009;6:60-75.
28. Kautzky-Willer A, Harreiter J, Pacini G. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. *Endocr Rev*. 2016;37(3):278-316.
29. Power ML, Schulkin J. Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *Br J Nutr*. 2008;99(5):931-940.
30. Di Angelantonio E, Sarwar N, Perry P, et al. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA*. 2009;302:1993-2000.
31. Sarwar N, Gao P, Seshasai S, et al. Emerging risk factors collaboration diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet*. 2010;375(9733):2215-2222.
32. Janghorbani M, Momeni F, Dehghani M. Hip circumference, height and risk of type 2 diabetes: systematic review and meta-analysis. *Obes Rev*. 2012;13(12):1172-1181.
33. Li K, Feng T, Wang L, et al. Causal associations of waist circumference and waist-to-hip ratio with type II diabetes mellitus: new evidence from Mendelian randomization. *Mol Genet Genomics*. 2021;296(3):605-613.
34. Feng J, He S, Chen X. Body adiposity index and body roundness index in identifying insulin resistance among adults without diabetes. *Am J Med Sci*. 2019;357(2):116-123.
35. de Oliveira CM, Pavani J, Krieger JE, de Oliveira AR, Mourão-Junior CA, da Costa PA. Body adiposity index in assessing the risk of type 2 diabetes mellitus development: the Baependi heart study. *Diabetol Metab Syndr*. 2019;11:76.
36. Schulze MB, Thorand B, Fritsche A, et al. Body adiposity index, body fat content and incidence of type 2 diabetes. *Diabetologia*. 2012;55(6):1660-1667.
37. Bellary S, Kyrou I, Brown JE, Bailey CJ. Type 2 diabetes mellitus in older adults: clinical considerations and management. *Nat Rev Endocrinol*. 2021;17(9):534-548.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Saberi-Karimian M, Mansoori A, Bajgiran MM, et al. Data mining approaches for type 2 diabetes mellitus prediction using anthropometric measurements. *J Clin Lab Anal*. 2023;37:e24798. doi:10.1002/jcla.24798