Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 9, No. 2, 2021, 193-202.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Developing a Novel Continuous Metabolic Syndrome Score: A Data Mining Based Model

Mohsen Saffarian[1], Vahide Babaiyan[2*], Kokab Namakin[3], Fatemeh Taheri[3] and Toba Kazemi[4]

1. Department of Industrial Engineering, Birjand University of Technology, Birjand, Iran.
2. Department of Computer Engineering, Birjand University of Technology, Birjand, Iran.
3. Cardiovascular Diseases Research Centre, Department of Pediatric, Birjand University of Medical Sciences, Birjand, Iran.
4. Cardiovascular Diseases Research Centre, Department of Cardiology, Birjand University of Medical Sciences, Birjand, Iran.

Article Info.

Article History:
Received 12 January 2021
Revised 30 March 2021
Accepted 10 April 2021

DOI:10.22044/jadm.2021.10433.2179

Keywords:
Metabolic Syndrome, Linear Discriminant Analysis, Cardiovascular Risk Factors, Decision Tree Model.

**Corresponding author: babaiyan@birjandut.ac.ir (V. Babaiyan).*

Abstract

Today, metabolic syndrome in the age group of children and adolescents has become a global concern. In this work, a data mining model is used in order to determine a continuous metabolic syndrome (cMetS) score using the linear discriminate analysis (cMetS-LDA). The decision tree model is used to specify the calculated optimal cut-off point cMetS-LDA. In order to evaluate the method, the multi-layer perceptron neural network (NN) and the support vector machine (SVM) models are used, and the statistical significance of the results is tested with the Wilcoxon signed-rank test. According to the results of this test, the proposed CART is significantly better than the NN and SVM models. The ranking results in this work show that the most important risk factors in making cMetS-LDA are WC, SBP, HDL, and TG for the males, and WC, TG, HDL, and SBP for the females. Our research work results show that the high TG and central obesity have the greatest impacts on MetS, and that FBS has no effect on the final prognosis. The results obtained also indicate that in the preliminary stages of MetS, WC, HDL, and SBP are the most important influencing factors that play an important role in forecasting.

1. Introduction

Metabolic syndrome (MetS) is one of the most common metabolic disorders and a set of cardiovascular risk factors that cause chronic diseases such as CVD, diabetes mellitus, cancer, kidney disease, and mental illness [1-6]. MetS can also be described as a metabolic disorder associated with the atherosclerotic cardiovascular disease. According to the recommendations of the Adult Treatment Panel III, hypertension, abdominal obesity, hyperglycemia, and decreased serum concentration of high-density lipoprotein-cholesterol and hypertriglyceridemia have been introduced as the main components of MetS [7-9]. MetS in the age group of children and adolescents has become a global concern, especially in developing countries, where the cardiovascular mortality rates are 2.5 times higher, and the deaths are 1.5 times higher. This increases the cardiovascular mortality rate and the resulting mortality rate by 2.5 and 1.5 times, respectively [10]. In the recent decades, MetS has become a major concern in children and adolescents due to the factors such as the high prevalence of obesity, epidemiological transmission, double burden of eating disorders, and lifestyle changes. Childhood is when certain patterns such as lifestyle and diet are habitually formed. Although the symptoms of atherosclerotic disorders gradually become apparent in the later years, an early detection of the factors affecting it and modification of the lifestyle can significantly reduce the incidence of these disorders. Compared to the studies in adults,

there are no comprehensive studies on the age group of children and adolescents that examine the causes of MetS or its prevalence. Many studies have suggested the use of continuous scores instead of dichotomous scores in children and adolescents [11-14]. For example, according to [15], the use of cMetS reduces the duality factor, and is more sensitive and will lead to fewer errors than when using the metabolic syndrome stratified assessment. The statistical power is also increased when cMetS is used. The full details on calculating the cMetS score are published in [16].

Increasing MetS is gradually associated with increasing the level of each individual risk factor and increasing the number of risk factors. A continuous score is used as the best reflection of the progressive nature of this issue. There are several strategies for the construction of cMetS using its components.

In this work, a data mining model was used for this purpose using the linear discriminant analysis. Data mining is the process of finding the anomalies, patterns, and correlations in a large dataset to predict the results. Using a wide range of techniques, this information can be used in order to make more accurate predictions at higher speeds and with a greater reliability. The aim of this work was specifically to find the most important risk factors in the development of MetS, and also to evaluate the effectiveness of different models with data mining.

Machine learning has recently been shown to have a better performance over the traditional statistical modeling approaches [17-20]. Various approaches based on decision trees have been successfully applied to the medical data including the myocardial infarction [21], colon cancer [22], trauma [23], breast cancer [24], Alzheimer disease [25], cardiac surgery [26], and others [27-29]. However, the decision tree has its own strength and weakness. Therefore, comparing different algorithms can reduce the bias in the results, and can provide a more robust outcome.

In [30], MetS has been predicted by the machine learning models using the decision tree algorithms. They evaluated the accuracy of different types of decision tree algorithms that predict the status of MetS in the self-paying health examiners examined with FibroScan. In their results, Obesity, serum glutamic-oxalocetic transaminase, serum glutamic pyruvic transaminase, controlled attenuation parameter score, and glycated hemoglobin appeared as the significant risk factors in multivariate logistic regression. Although they used the machine learning techniques, they did not consider the cMetS score.

In contrast, [31, 32] determine the cMetS score in the age group of children but use the statistical analysis and do not use the machine learning techniques.

Referring to the strengths and weaknesses of the research works conducted in this field, we proposed a method that used the power of artificial intelligence in order to determine the continuous metabolic syndrome (cMetS) score in children and adolescents to achieve better results.

The rest of this paper is organized as what follows. The second part gives a complete description of the proposed method in details. The results and discussion are presented in the third section, and finally, in the fourth section, conclusions and future suggestions are expressed.

2. Proposed Method

In this section, we describe the proposed method step by step. According to the chart below, we first collected the data required for the current work. In the next steps, the statistical analysis and modeling are performed. It is noteworthy that the operations such as data balancing in datasets that are inherently unbalanced as well as the extraction of distinctive and fundamental features will have a significant impact on the model and the final result. Therefore, the pre-processing steps such as data balancing and feature extraction were performed before classification. Then using the CART method, we predicted the continuous and stratified target variables, and we also applied the multilayer perceptron neural network and support vector machine in order to evaluate the CART method.

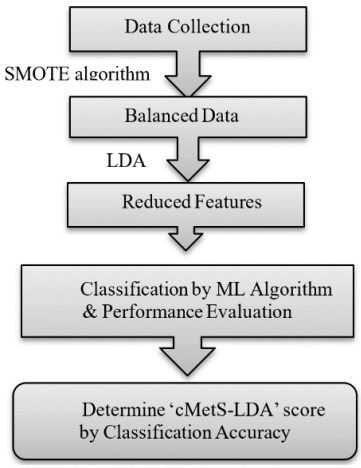

Figure 1. Proposed model based on the LDA dimensionality reduction technique.

2.1. Data Collection

In this descriptive cross-sectional research, 4340 students aged 6 to 18 years old in Birjand including 2,329 females and 2,011 males were studied in the years 2012 and 2013. The samples were chosen through multi-stage cluster sampling from the school-aged students.

The weight and height measurements were performed by the standard methods. The height was measured with an accuracy of 0.5 cm and the weight with a digital scale (Seca, German) with 100 g error. The waist circumference (WC) was measured with an accuracy of 0.5 cm in standing position at the distance between the last rib and the iliac in exhale. The blood pressure was measured under the standard circumstances.

The laboratory blood tests such as high-density lipoprotein (HDL), fasting blood sugar (FBS), and triglyceride (TG) were performed for the students. The blood samples were collected in 5 mL vacuum tubes (with separator gel, clot activator, Bacton Dickinson UK). The blood samples were separated at a distance of less than 15 min for 10 minutes' centrifugation by sigma centrifugation in 3000 RPM, and within fewer than an h, FBS, TG, HDL-C, LDL-C and total serum cholesterol (TC) were measured by enzymatic reagents (fully automated ROCHE COBAS INTEGRA analysis system). ATP III is a diagnostic criterion for MetS in this work. According to ATP III, at least three of the followings were defined as MetS: WC ≥ 90th percentile for age and gender, abdominal obesity, HDL < 40, TG ≥ 110 mg/dL, FBS ≥ 110 mg/dL, and diastolic or systolic blood pressure (SBP) ≥ 90th percentile. The exclusion criteria were genetic syndrome, endocrine disorders, physical impairments, and using drugs, which somehow affected the MetS components. From the population, we excluded the individuals with missing data on important variables (n = 521). The final population with size of 3819 students (242 abnormal and 3577 normal) was used in this work. The basic characteristics of the studied population are shown in Table 1.

Also the distribution of the studied population by gender and age is shown in Table 2.

The research dataset was partitioned into two subsets using the stratified random sampling: a training dataset consisted of 80% of the dataset, which was used for model building, and the remaining 20% was used for testing the model.

2.2. Data Balancing

Most prevalent classification models perform well when the classes of target variable are distributed identically, and the model challenges increase when the distribution of observations is imbalance [33]. Often the medical data is mainly combined by the majority normal class with only a minority of the abnormal class [34].

Table 1. Baseline characteristics of Birjand MetS study.

Field	Abnormal (242)	Normal (3577)	P-Value*
Weight	62.022**	41.035	1
Length	155.017	146.791	1
WC	81.153	63.586	1
SBP	121.591	101.779	1
DBP	68.967	57.897	1
FBS	92.033	88.384	1
CHOL	171.905	157.067	1
TG	157.269	82.642	1
HDL	36.64	49.934	1
LDL	101.9	86.609	1
BMI	25.289	18.44	1

*P-Value of significant test for difference between mean of variables for abnormal and normal groups. P-Values greater than 0.95 indicates an important and significant difference

** Mean of variable

Table 2. Distribution of studied population by gender and age.

Distribution of dataset		Abnormal	Normal
Gender	**Male**	139	1600
	Female	103	1977
Age group	**6-11**	92	1564
	12-18	150	2013

In this case, the standard and popular models tend to generate a high precision for the majority class and a low precision for the minority. The oversampling techniques are used in order to dominate this problem by adjusting the prior probability of the majority and minority class in the training dataset to acquire more balanced data in each class. Some studies in the medical science have shown the effectiveness of the SMOTE algorithm to solve the problem of unbalanced data [34, 35]. This algorithm helps to balance the data by producing artificial samples of the minority class data. Some studies in the medical science have shown the effectiveness of the SMOTE algorithm to solve the problem of unbalanced data [34, 35]. In the present work, we balanced the training datasets using SMOTE.

2.3. Linear Discriminant Analysis (LDA)

LDA is a type of supervised feature extractor that has been widely used [36-39]. Its goal is to transform the original feature space into the most discriminate one. In order to achieve this goal, LDA seeks to reduce the ratio of the inter-class scattering matrix to the intra-class scattering

matrix in the dataset [40]. The two scattering matrices mentioned in (1) and (2) are introduced, respectively.

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \quad (1)$$

$$S_b = \sum_{j=1}^{c} (\mu_i - \mu)(\mu_i - \mu)^T \quad (2)$$

Where c and N_j denote the number of classes and the number of instances in class j, respectively, x_i^j denotes the ith instance of class j, μ_j is the average of class j, and μ represents the average of all classes. One way to obtain the LDA's goal is to maximize the ratio $\frac{\det|S_b|}{\det|S_w|}$ [40].

The eigenvectors of $S_w^{(-1)}S_b$ are the orientations of the new feature space.

2.4. Classification Modeling and Performance Evaluation

The CART method is a popular DT model [41], and is used to predict the continuous and categorical target variables. CART splits the data subsets using all variables for creating two child nodes. The best variable is found using an impurity measure to produce as homogenous as possible data subsets with respect to the target variable. The CART procedure is terminated by cost-complexity tree pruning. In this step, in order to evaluate the efficiency of the CART method, the multi-layer perceptron neural network and the support vector machine are used in comparisons.

3. Results

3.1. Performance Evaluation

In the first stage, we implemented the LDA method to obtain cMetS-LDA for the males and females. The measures obtained for the males and females are shown in (3) and (4), respectively. The CART algorithm is used to specify the best cMetS-LDA cut-off point to forecast MetS. Figure 2 shows the best cut-point of cMetS-LDA obtained using the CART model that for the males and females are 0.603 and 0.308, respectively. The CART models are constructed using the balanced training subsets of the male and female students.

$$\begin{aligned} cMetS-LDA(Males) = \\ -0.12WEIGHT - 0.38LENGHT \\ +0.77WC + 0.51SBP \\ -0.014DBP + 0.026FBS \\ +0.359CHOL + 0.33TG \\ -0.5HDL - 0.21LDL \end{aligned} \quad (3)$$

$$\begin{aligned} cMetS-LDA(Females) = \\ -0.12WEIGHT - 0.65LENGHT \\ +0.73WC + 0.44SBP \\ -0.093DBP + 0.15FBS \\ -0.52CHOL + 0.5TG \\ -0.42HDL - 0.5LDL \end{aligned} \quad (4)$$

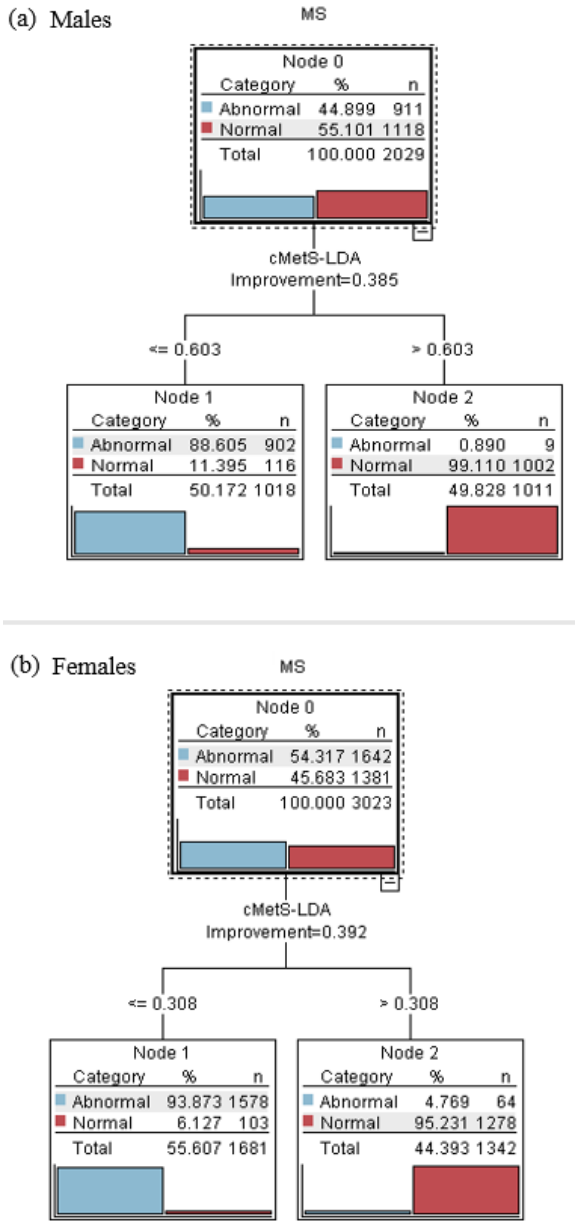

Figure 2. CART model for (a) males and (b) females.

Figure 3 displays the histogram of the predicted MetS using the obtained cut-points.

According to Figure 3, the constructed score by LDA splits the instances in two classes into two nearly separated intervals. Thus using cMetS-

LDA, the classifier can easily classify a new instance x by executing the simple rule that for males this rule is "If cMet-LDA ≤ 0.603. Then the class is abnormal (with a probability of 0.88), and If cMet-LDA > 0.603, then the class is normal (with a probability of 0.99)".

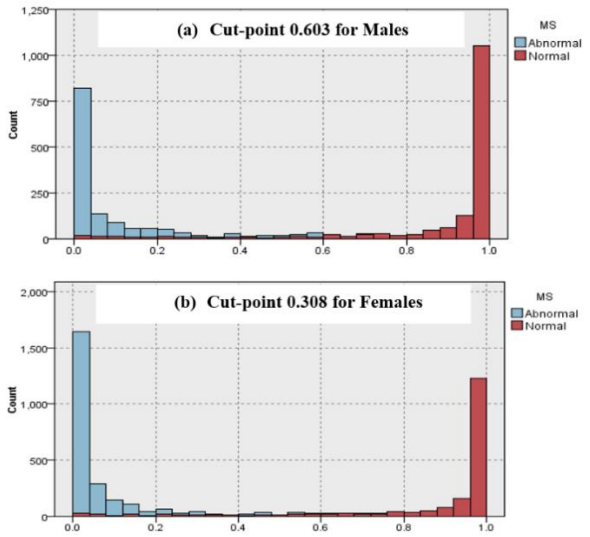

Figure 3. Histogram of the predicted MS using the obtained cut-points for (a) males and (b) females.

The ROC plot for the models on the testing datasets is shown in Figure 4.

In addition to the CART method, in this classification problem, we used two well-known models, the multi-layered perceptron neural network and the support vector machine, as a criterion to ensure the superiority of the proposed method by conducting the experiments.

The suitable results were obtained for a single-layered neural network with 8 hidden neurons with hyperbolic-tangent activation function and SVM with radial basis function (RBF) kernel functions. The performance measures for the proposed models are shown in Table 3. The accuracy measures of the CART model obtained from cMetS-LDA in Table 3 and its ROC curve shows that using the obtained cut-points has a high accuracy in predicting MS.

In order to further evaluate the performance and also test the statistical significance of the results of the CART model, a Wilcoxon signed-rank test [42] was carried out on the mean results of the proposed models shown in Table 3.

In order to conduct this test, we supposed the following hypotheses:

- H0: There is no difference among the performance of CART and other models.
- H1: A difference exists among the performance of CART and the other models.

The results of the Wilcoxon signed-rank test are shown in Table 4.

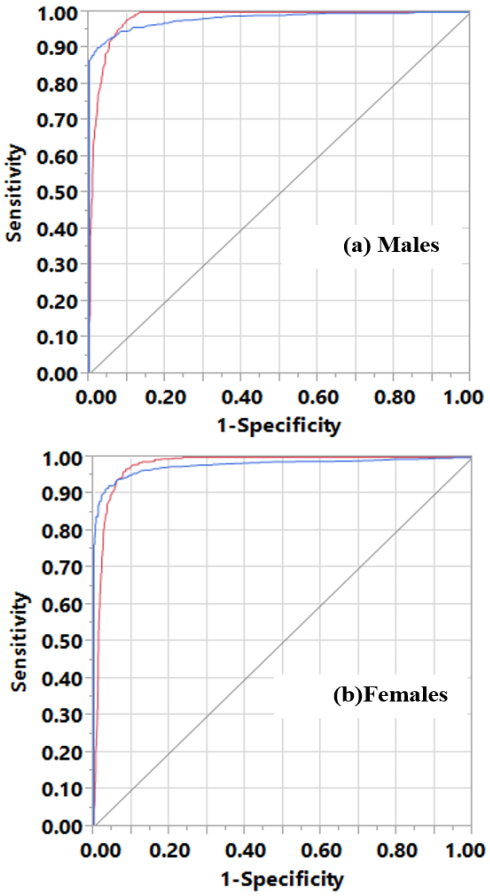

Figure 4. ROC curve of the obtained tree using cMetS-LDA for metabolic syndrome for (a) males and (b) females.

Table 3. Performances of the proposed models for males and females.

	Males			Females		
	CART	**SVM**	**NN**	**CART**	**SVM**	**NN**
Sensitivity	0.966	0.921	0.953	0.97	0.951	0.935
specificity	0.919	0.852	0.901	0.893	0.91	0.882
Accuracy	0.936	0.913	0.925	0.921	0.908	0.897
G-Measure	0.942	0.93	0.915	0.931	0.893	0.915

With regard to Table 4, the proposed CART is significantly better than the NN and SVM models. Determination of the key variables is a vital requirement in many applications. In this research work, we saw that the feature selection was one of the main issues in various fields. In constructing the cMetS-LDA score, all the risk factors were graded in order to determine the predictors of MetS based on their significance. Among all the factors for predicting MetS, some of them affect the accuracy of the proposed model more than the others. In this work, we utilized a heuristic method proposed by Johnson [43] for ranking the risk factors in constructing cMetS-LDA for the

males and females. The results of this method in the ranking of the risk factors are shown in Figures 5 and 6, respectively.

Table 4. Results of Wilcoxon signed-rank test for all compared models

Moddel	Average of best results	Std-dev	Base model for comparison	Z value	Pr (> |Z|)
SVM	0.898476	0.302071	CCMBO	7.123991	0.0
NN	0.909828	0.286474		5.122593	0.0

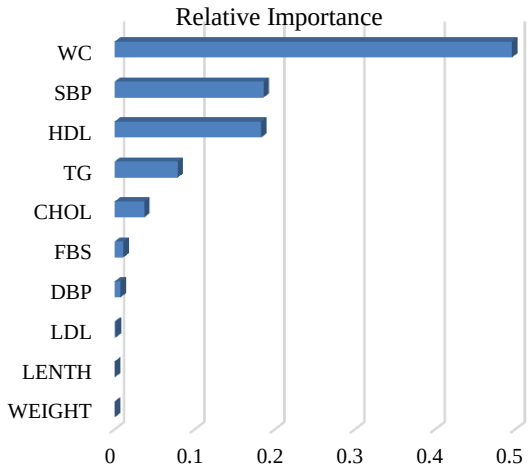

Figure 5. Ranking of risk factors in cMetS-LDA constructing for males.

3.2. Discussion

The metabolic syndrome has a significant effect on the onset and development of type 2 diabetes and cardiovascular diseases (CVDs). It is currently a serious social health-related problem, and may occur at any age, although in a varying degree [7]. Its main variables are well-defined in adults, while there is no clear global description of MetS in the age group for children and adolescents [44]. There is a different prevalence among the MetS studies in children [45-49]. Therefore, we require a larger dataset to examine the relationship between the risk factors. Accordingly, in many studies in the field of pediatrics, by adapting the cut-off points of each metabolic parameter for children and adolescents, this definition has been used for the adults. In various studies, the effect of the MetS components from childhood to adulthood has been clearly demonstrated [50].

Due to the occurrence of obesity and concomitant diseases in children, a primary prevention of atherosclerosis is provided by screening children at the metabolic risk, which is very important [51, 52]. All the MetS risk components are considered by a continuous variable. It is assumed that all the influential components in MetS are also important and effective in CVD.

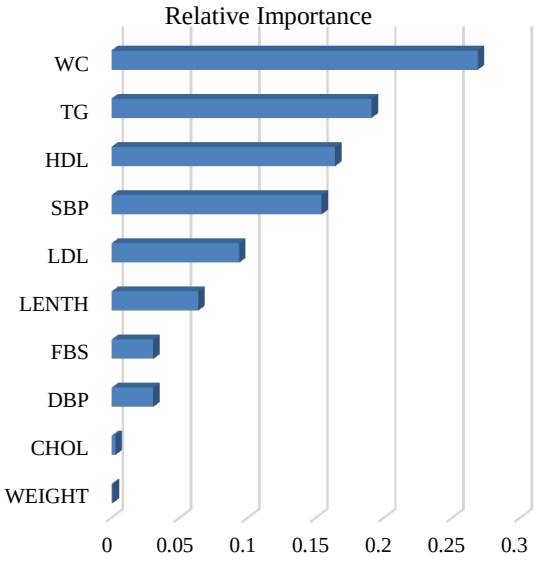

Figure 6. Ranking of risk factors in cMetS-LDA constructing for females.

The recent cross-sectional research reports indicate the effectiveness of cMetS because it is statistically more accurate and more sensitive than the binary classification [12, 53]. Given that in many children there is only one or two influential risk factors, the present work shows the importance of predicting the risk of atherosclerosis in children using cMetS. Various studies have forecasted cMetS with principal component analysis (PCA) and z-score methods, and it has been carefully evaluated with both the ROC curves and the below curve area. So far, no study has been carried out with the LDA method. The best cut-off point for predicting cMets-LDA was obtained using the CART method, and was estimated 0.603 (sensitivity 0.97, specificity 0.91) in the males and 0.308 (sensitivity, 0.97; specificity, 0.89) in the females with the below curve area of AUC: 0.94 in the males and AUC: 0.92 in the females. These results are more compared to Shi *et al*. (2015) as the below curve area for cMetS-Z and cMets-PCA, as a predictor of MetS, in their study have been reported to be 0.93 and 0.91, respectively [11], and therefore, our results have a higher accuracy. To the contrary, Eisenmann *et al*. (2010) have reported a higher accuracy than us (cut-off point: 3.72, sensitivity: 100%, specificity: 93.9%, AUC: 0.978) [44]. These results have been reported by Shafiee *et al*. (2013) as follows (optimal cut-ponit for cMetS: 2.93, sensitivity: 92%, specificity:

91%, AUC: 0.96), which was calculated to be 2.97 for the males and 3.26 for the females [50].

In our previous studies, the incidence and prevalence of MetS in the age groups 11-18 years and 18-13 years were 5.3% and 6.9%, respectively [7, 45]. Overall, 53.5% of the samples were normal and 46.5% had at least one abnormal component of MetS. Among the risk factors influencing MetS forecasting, some of them are more important in the accuracy of the proposed cMetS-LDA model. In this work, the waist circumference (WC) was an important predictor in men and women. In the cMetS-LDA model, the important risk for the males was WC, SBP, HDL, and TG, and for the females, was WC, TG, HDL, and SBP, respectively. The visceral fat increases inflammation and oxidative stress due to the release of inflammatory mediators, so it is considered a risk factor for MetS and other metabolic complications.

These conditions may have biological effects affecting the peripheral metabolic, vascular, and endocrine processes. The process of low-grade inflammation may predispose to the metabolic risk factors and exacerbate the consequences. In addition, WC and increased visceral fat are two powerful and independent predictors of the metabolic changes [54-56]. Despite the difficulty comparing the present work with the other ones, the results obtained showed central adiposity reflects on the health of children. Dyslipidemia is associated with the MetS and cardiovascular complications, and our results suggest that the study population is at a higher risk in the later years. According to several studies [50, 52] as well as the results of our study, the central obesity and high TG have the most impact on MetS and cMetS score. In this work, FBS was not an effective factor in predicting the outcome. Our work shows that the preliminary stages of MetS, WC, HDL-c, and SBP are the effective components. Given the low MetS outbreak in children, conducting the communication studies using the binary variable is challenging. Thus we used the continuous cMets score as a more sensitive criterion. The large sample size and homogeneity of the studied samples are the strengths of this work. According to our latest knowledge, there are no previous studies based on the cMetS-LDA model performed on children.

4. Conclusions

In this work, a data mining model was utilized in order to determine the score of a continuous metabolic syndrome (cMeS) based on its effective components using the linear discriminant analysis (cMetS-LDA). The decision tree modeling was utilized in order to determine the optimum cut-off point of the obtained cMetS-LDA. In order to evaluate the CARD method, a comparison with the NN and SVM models was performed, and the superiority of the decision tree in the results was shown. We also tested the statistical significance of the results of the CART model.

A Wilcoxon signed-rank test was carried out on the mean results of the proposed models. With regard to the results of this test, the proposed CART is significantly better than the NN and SVM models.

The convergent validity for the cMetS-LDA model was shown in this work that is becoming widely used in the pediatric epidemiological research works.

Our results suggest that the central obesity and high TG have the greatest effect on the MetS score. In this work, FBS was not an effective factor in predicting the outcome. Our study shows that the preliminary stages of MetS, WC, HDL, and SBP are the main components. We suggest that more research works should be done in order to determine the cut-off points for the metabolic risk factors in children, which include the waist circumference measurements, as these measurements may help the physician to intervene immediately.

References

[1] A. Pan, N. Keum, O. I. Okereke, Q. Sun, M. Kivimaki, R. R. Rubin, and F. B. Hu, "Bidirectional association between depression and metabolic syndrome: a systematic review and meta-analysis of epidemiological studies," *Diabetes care*, vol. 35, no. 5, pp. 1171-1180, 2012.

[2] D. Lent-Schochet, M. McLaughlin, N. Ramakrishnan, and I. Jialal, "Exploratory metabolomics of metabolic syndrome: A status report," *World journal of diabetes*, vol. 10, no. 1, pp. 23, 2019.

[3] W. Borena, M. Edlinger, T. Bjørge, C. Häggström, B. Lindkvist, G. Nagel, A. Engeland, T. Stocks, S. Strohmaier, and J. Manjer, "A prospective study on metabolic risk factors and gallbladder cancer in the metabolic syndrome and cancer (Me-Can) collaborative study," *PloS one*, vol. 9, no. 2, pp. e89368, 2014.

[4] A. Maleki, M. Montazeri, N. Rashidi, M. Montazeri, and E. Yousefi-Abdolmaleki, "Metabolic syndrome and its components associated with chronic kidney disease," *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, vol. 20, no. 5, pp. 465, 2015.

[5] M. Jari, M. Qorbani, M. E. Motlagh, R. Heshmat, G. Ardalan, and R. Kelishadi, "Association of overweight and obesity with mental distress in Iranian

adolescents: the CASPIAN-III study," *International journal of preventive medicine*, vol. 5, no. 3, pp. 256, 2014.

[6] M. K. Birarra, and D. A. Gelayee, "Metabolic syndrome among type 2 diabetic patients in Ethiopia: a cross-sectional study," *BMC cardiovascular disorders*, vol. 18, no. 1, pp. 1-12, 2018.

[7] M. Zardast, K. Namakin, T. Chahkandi , F. Taheri , T. Kazemi , and B. Bijari, "Prevalence of metabolic syndrome in elementary school children in East of Iran," *J Cardiovasc Thorac Res*, vol. 7, no. 4, pp. 158-163, 2015.

[8] M.-K. Lee, K. Han, M. K. Kim, E. S. Koh, E. S. Kim, G. E. Nam, and H.-S. Kwon, "Changes in metabolic syndrome and its components and the risk of type 2 diabetes: a nationwide cohort study," *Scientific reports*, vol. 10, no. 1, pp. 1-8, 2020.

[9] P. Golabi, M. Otgonsuren, L. de Avila, M. Sayiner, N. Rafiq, and Z. M. Younossi, "Components of metabolic syndrome increase the risk of mortality in nonalcoholic fatty liver disease (NAFLD)," *Medicine*, vol. 97, no. 13, 2018.

[10] A. N. Rodrigues, G. R. Abreu, R. S. Resende, W. L. Goncalves, and S. A. Gouvea, "Cardiovascular risk factor investigation: a pediatric issue," *International journal of general medicine*, vol. 6, pp. 57, 2013.

[11] P. Shi, J. M. Goodson, M. L. Hartman, H. Hasturk, T. Yaskell, and J. Vargas, "Continuous Metabolic Syndrome Scores for Children Using Salivary Biomarkers," *PLoS ONE*, vol. 10, no. 9, *pp. e0138979, 2015*.

[12] D. Pandit , S. Chiplonkar, A. Khadilkar, A. Kinare, and V. Khadilkar "Efficacy of a continuous metabolic syndrome score in Indian children for detecting subclinical atherosclerotic risk," *Int J Obes (Lond), vol. 35, no. 10, pp. 1318-1324, 2011*.

[13] H.-S. Ejtahed, M. Qorbani, M. E. Motlagh, P. Angoorani, S. Hasani-Ranjbar, H. Ziaodini, M. Taheri, Z. Ahadi, S. Beshtar, and T. Aminaee, "Association of anthropometric indices with continuous metabolic syndrome in children and adolescents: the CASPIAN-V study," *Eating and Weight Disorders-Studies on Anorexia*, Bulimia and Obesity, vol. 23, no. 5, pp. 597-604, 2018.

[14] S. P. Sawant, and A. S. Amin, "Use of continuous metabolic syndrome score in overweight and obese children," *The Indian Journal of Pediatrics*, vol. 86, no. 10, pp. 909-914, 2019.

[15] S. Rose, F. F. Dieny, and A. Tsani, "The Correlation between Waist-to-Height Ratio (WHtR) and Second to Fourth Digit Ratio (2D: 4D) with an Increase in Metabolic Syndrome Scores in Obese Adolescent Girls," *Electronic Journal of General Medicine*, vol. 17, no. 3, 2020.

[16] G. Shafiee, R. Kelishadi, R. Heshmat, M. Qorbani, M. E. Motlagh, T. Aminaee, G. Ardalan, M. Taslimi, P. Poursafa, and B. Larijani, "First report on the validity of a continuous Metabolic Syndrome score as an indicator for Metabolic Syndrome in a national sample of paediatric population—the CASPIAN-III study," *Endokrynologia Polska*, vol. 64, no. 4, pp. 278-284, 2013.

[17] A. L. Beam, and I. S. Kohane, "Big Data and Machine Learning in Health Care," *JAMA*, vol. 319, no. 13, pp. 1317-1318, 2018.

[18] J. H. Chen, and S. M. Asch, "Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations," *New England Journal of Medicine*, vol. 376, no. 26, pp. 2507-2509, 2017.

[19] B. A. Goldstein, A. M. Navar, and R. E. Carter, "Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges," *European* Heart Journal, vol. 38, no. 23, pp. 1805-1814, 2016.

[20] S. M. Cho, P. C. Austin, H. J. Ross, H. Abdel-Qadir, D. Chicco, G. Tomlinson, C. Taheri, F. Foroutan, P. R. Lawler, and F. Billia, "Machine Learning Compared To Conventional Statistical Models For Predicting Myocardial Infarction Readmission And Mortality: A Systematic Review," *Canadian Journal of Cardiology*, 2021.

[21] H. C. Lee, J. S. Park, J. C. Choe, J. H. Ahn, H. W. Lee, J.-H. Oh, J. H. Choi, K. S. Cha, T. J. Hong, and M. H. Jeong, "Prediction of 1-Year Mortality from Acute Myocardial Infarction Using Machine Learning," *The American Journal of Cardiology*, vol. 133, pp. 23-31, 2020.

[22] M. V. Bhargavi, V. R. Mudunuru, and S. Veeramachaneni, "Colon cancer stage classification using decision trees," Data Engineering and Communication Technology, pp. 599-609: Springer, 2020.

[23] A. G. Aydin, O. Eray, A. V. Sayrac, A. Oskay, And U. D. Ulusar, "The Reliability of an Artificial Intelligence Tool,'Decision Trees', in Emergency Medicine Triage," 2020.

[24] K. T. Win, "Comparison of C4. 5 and Weighted C4. 5 Decision Trees for Breast Cancer Classification," Unversity of Computer Studies, Yangon, 2020.

[25] G. Battineni, N. Chintalapudi, and F. Amenta, "Late-Life Alzheimer's Disease (AD) Detection Using Pruned Decision Trees," *Int J Brain Disord Treat*, vol. 6, pp. 033, 2020.

[26] M. Mamprin, S. Zinger, P. de With, J. Zelis, and P. Tonino, "Gradient boosting on decision trees for mortality prediction in transcatheter aortic valve implantation." pp. 325-329.

[27] M. M. Ghiasi, "Implementing decision tree-based algorithms in medical diagnostic decision support systems," Memorial University of Newfoundland, 2020.

[28] A. Alaoui, and Z. Elberrichi, "Enhanced Ant Colony Algorithm for Best Features Selection for a Decision Tree Classification of Medical Data," Critical Approaches to Information Retrieval Research, pp. 278-293: IGI Global, 2020.

[29] Y. Wei, X. Wang, and M. Li, "Intelligent Medical Auxiliary Diagnosis Algorithm Based on Improved Decision Tree," *Journal of Electrical and Computer Engineering*, vol. 2020, 2020.

[30] C.-S. Yu, Y.-J. Lin, C.-H. Lin, S.-T. Wang, S.-Y. Lin, S. H. Lin, J. L. Wu, and S.-S. Chang, "Predicting metabolic syndrome with machine learning models using a decision tree algorithm: Retrospective cohort study," *JMIR medical informatics*, vol. 8, no. 3, pp. e17110, 2020.

[31] J. C. Eisenmann, K. R. Laurson, K. D. DuBose, B. K. Smith, and J. E. Donnelly, "Construct validity of a continuous metabolic syndrome score in children," *Diabetology & metabolic syndrome*, vol. 2, no. 1, pp. 1-8, 2010.

[32] H.-S. Ejtahed, Z. Mahmoodi, M. Qorbani, P. Angoorani, M. E. Motlagh, S. Hasani-Ranjbar, H. Ziaodini, M. Taheri, R. Heshmat, and R. Kelishadi, "A comparison between body mass index and waist circumference for identifying continuous metabolic syndrome risk score components in Iranian school-aged children using a structural equation modeling approach: the CASPIAN-V study," *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, pp. 1-8, 2020.

[33] N. V. Chawla, "SMOTEBoost: Improving prediction of the minority class in boosting, in Knowledge Discovery in Databases: PKDD " *Springer*, vol. P.107-119, 2003.

[34] A. Ramezankhani, "The impact of oversampling with SMOTE on the performance of 3 classifiers in prediction of type 2 diabetes," *Medical Decision Making*, vol. p. 0272989X14560647., 2014.

[35] N. Chawla, "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. p. 321-357, 2002.

[36] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of eugenics*, vol. 8, no. 4, pp. 376-386, 1938.

[37] K. Fukunaga, Introduction to statistical pattern recognition: Elsevier, 2013.

[38] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI communications*, vol. 30, no. 2, pp. 169-190, 2017.

[39] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54776-54788, 2020.

[40] A. M. Martínez, and A. C. Kak, "Pca versus lda," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 2, pp. p. 228-233, 2001.

[41] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, Classification and regression trees: CRC press, 1984.

[42] S. García, D. Molina, M. Lozano, and F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization," J*ournal of Heuristics*, vol. 15, no. 6, pp. 617-644, 2009.

[43] J. W. Johnson, "A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression," *Multivariate Behav Res*, vol. 35, no. 1, pp. p. 1-19, 2000.

[44] J. C. Eisenmann, K. R. Laurson, K. D. DuBose, B. K. Smith, and J. E. Donnelly, "Construct validity of a continuous metabolic syndrome score in children.," *Diabetology & metabolic syndrome*, vol. 2, no. 1, pp. 8, 2010.

[45] F. Taheri , K. Namakin, M. Zardast, T. Chahkandi , T. Kazemi , and B. Bijari, " Cardiovascular Risk Factors: A Study on the Prevalence of MS among 11-18 Years Old School Children in East of Iran, 2012," Nutr Food SCI Res, vol. 2, no. 1, pp. 27-34, 2015.

[46] B. Battaloglu Inanc, "Metabolic Syndrome in School Children in Mardin, South-Eastern of Turkey," Eurasian J Med, vol. 46, no. 3, pp. 156-163, 2014.

[47] W. Ahrens, L. A. Moreno, S. Mårild, D. Molnár, A. Siani, and S. De Henauw, "Metabolic syndrome in young children: definitions and results of the IDEFICS study," *Int J Obes (Lond)*, vol. 38, no. Suppl 2, pp. S4-s14, 2014.

[48] N. Rizk, M. Amin, and M. Yousef, "A pilot study on metabolic syndrome and its associated features among Qatari school children," *Int J Gen Med*, vol. 4, pp. 521-525, 2011.

[49] A. Z. Iqbal, S. Basharat , and A. Basharat "Prevalence of the metabolic syndrome and its component abnormalities among school age Pakistani children," *J Ayub Med Coll Abbottabad*, vol. 26, no. 2, pp. 194-199, 2014.

[50] G. Shafiee, R. Kelishadi , R. Heshmat, M. Qorbani, M. E. Motlagh , and T. Aminaee, "First report on the validity of a continuous Metabolic Syndrome score as an indicator for Metabolic Syndrome in a national sample of paediatric population—the CASPIAN-III study," *Endokrynol Pol*, vol. 64, no. 4, pp. 278-284, 2013.

[51] F. Costa Teixeira, F. E. Felix Pereira , A. Fernandes Pereira , and B. Gonçalves Ribeiro "Metabolic syndrome's risk factors and its association with nutritional status in school children," *Prev Med* Rep, vol. 6, no. 2017, pp. 27-32, 2017.

[52] C. Cadenas-Sanchez , J. R. Ruiz, I. Labayen, I. Huybrechts, Y. Manios, and M. González-Gross, "Prevalence of Metabolically Healthy but Overweight/Obese Phenotype and Its Association With Sedentary Time, Physical Activity, and Fitness.," *J* Adolesc Health., vol. 61, no. 1, pp. 107-114, 2017.

[53] I. S. Okosun, J. M. Boltri, R. Lyn, and M. Davis-Smith "Continuous metabolic syndrome risk score, body mass index percentile, and leisure time physical activity in American children," *J Clin Hypertens (Greenwich)*, vol. 1, no. 2, pp. 636-644, 2010.

[54] T. Cohen, T. Hazell, C. A. Vanstone, C. Rodd, and H. A. Weiler, "A family-centered lifestyle intervention for obese six- to eight-year-old children: results from a one-year randomized controlled trial conducted in Montreal, Canada," *Can J Public Health*, vol. 107, no. 4-5, pp. 453-460, 2016.

[55] R. M. Mantovani, N. P. Rocha, D. M. Magalhães, I. G. Barbosa , A. L. Teixeira, and A. C. Simões E Silva "Early changes in adipokines from overweight to obesity in children and adolescents," *J Pediatr (Rio J)*, vol. 92, no. 6, pp. 624-630, 2016.

[56] S. M. Barbalho, M. Oshiiwa, F. C. Sato Fontana, E. Ribeiro Finalli , M. Paiva Filho , and A. P. Machado Spada, "Metabolic syndrome and atherogenic indices in school children: A worrying panorama in Brazi," D*iabetes Metab Syndr*, vol. S1871-4021, no. 17, pp. 30003-30006, 2017.

توسعه یک سیستم امتیازدهی پیوسته برای سندرم متابولیک: مدل مبتنی بر داده کاوی

محسن صفاریان[۱]، وحیده بابائیان[۲،*] ، کوکب نمکین[۳]، فاطمه طاهری[۳] و طوبی کاظمی[۴]

[۱] گروه مهندسی صنایع، دانشگاه صنعتی بیرجند، بیرجند، ایران.

[۲] گروه مهندسی کامپیوتر، دانشگاه صنعتی بیرجند، بیرجند، ایران.

[۳] مرکز تحقیقات بیماری‌های قلب و عروق، گروه کودکان، دانشگاه علوم پزشکی بیرجند، بیرجند، ایران.

[۴] مرکز تحقیقات بیماری‌های قلب عروق، گروه قلب، دانشگاه علوم پزشکی بیرجند، بیرجند، ایران.

ارسال ۲۰۲۱/۰۱/۱۲؛ بازنگری ۲۰۲۱/۰۳/۳۰؛ پذیرش ۲۰۲۱/۰۴/۱۰

چکیده:

امروزه سندرم متابولیک در گروه سنی کودکان و نوجوانان به یک نگرانی جهانی تبدیل شده است. در این مقاله، از یک مدل داده کاوی در امتیازدهی پیوسته سندرم متابولیک با استفاده از آنالیز تشخیصی خطی (cMetS-LDA) استفاده شده است. مدل درخت تصمیم در تعیین نقطه قطع بهینه‌ی امتیازدهی پیوسته سندرم متابولیک بکار گرفته شده است. دو مدل شبکه عصبی پرسپترون چندلایه و ماشین بردار پشتیبان با مدل پیشنهادی، مورد مقایسه قرار گرفته اند و از نظر آماری نتایج با آزمون Wilcoxon نیز ارزیابی شده اند. طبق نتایج این آزمون، مدل پیشنهادی به طور قابل توجهی بهتر از دو مدل شبکه عصبی پرسپترون چندلایه و ماشین بردار پشتیبان عمل نموده است. نتایج رتبه بندی در این مطالعه نشان می‌دهدکه مهمترین عوامل خطر در تعیین امتیاز پیوسته سندرم متابولیک، دورکمر[۱] ، فشارخون سیستولیک[۲]، لیپوپروتئین با چگالی بالا [۳]و تری‌گلیسرید[۴] برای مردان و دورکمر، تری گلیسرید، لیپوپروتئین با چگالی بالا و فشارخون سیستولیک برای زنان بودند. نتایج تحقیقات ما نشان می دهد که تری‌گلیسرید بالا و چاقی مرکزی بیشترین تأثیر را در سندروم متابولیک دارند و قند خون ناشتا[۵] تأثیری در پیش بینی نهایی ندارد. همچنین نتایج نشان می دهد که در مراحل اولیه سندروم متابولیک، دور کمر، لیپوپروتئین با چگالی بالا و فشارخون سیستولیک مهمترین عوامل تأثیرگذار هستند که نقش مهمی در پیش‌بینی دارند.

کلمات کلیدی: سندروم متابولیک، آنالیز تشخیصی خطی، عوامل خطر قلبی و عروقی، مدل درخت تصمیم گیری.

۱ WC
۲ SBP
۳ HDL
۴ TG
۵ FBS