Assessment of Anthropometric Measurements for Obesity and Abnormal Body Fat Percentage Diagnosis Using k-means as Clustering Technique

Alexandra La Cruz[1(✉)], Erika Severeyn[2], Jesús Velásquez[2], Héctor Herrera[3], and Sara Wong[4]

[1] Faculty of Engineering, Ibagué University, Ibagué, Tolima, Colombia
alexandra.lacruz@unibague.edu.co

[2] Department of Thermodynamics and Transfer Phenomena, Simón Bolívar University, Caracas, Venezuela
{severeynerika,jmvelasquezf}@usb.ve

[3] Department of Biological and Biochemical Process Technology, Simón Bolívar University, Caracas, Venezuela
haherrera@usb.ve

[4] Department of Electronics and Circuits, Simón Bolívar University, Caracas, Venezuela
swong@usb.ve

Abstract. The increased prevalence of overweight and obesity has become a major factor in public spending in countries around the world. The diagnosis of overweight and obesity is based on body mass index (BMI) and body fat percentage (BFP). The World Health Organization proposed BMI cut-off points to define overweight and obesity. Recently epidemiological studies established as normal BFP a BFP < 25 for men and BFP < 30 for women. A high correlation between a high BMI, abnormal BFP and skin thinness have been found in numerous studies. The aim of this work is to evaluate the k-means clustering algorithm using anthropometric measurements for the classification of subjects with overweight/obesity and abnormal BFP. Precision (P), accuracy (Acc) and recall (R) were calculated to evaluate the efficiency of the method to classify overweight/obesity and abnormal BFP. Results of this research suggest that the k-means method applied to anthropometric measurements can make an acceptable classification of overweight/obesity and abnormal BFP. The arm circumferences values show the best Acc, P and R (0.79, 0.84 and 0.71) compared to all other measurements for overweight/obesity diagnosis, otherwise, suprailiac and abdominal skinfolds values show the best Acc, P and R (0.73, 0.73 and 0.64) compared to all other measurements for abnormal BFP diagnosis. Results that are supported by studies asserting a strong relationship between arm circumferences, abdominal skinfold, suprailiac skinfold, BFP and BMI. Other machine learning techniques, such as neural networks and the support vector machine, will be studied in the future to assess the relationship between BMI, BFP and anthropometric measurements.

G. Rodriguez Morales et al. (Eds.): TICEC 2020, CCIS 1307, pp. 177–191, 2020.
https://doi.org/10.1007/978-3-030-62833-8_15

Keywords: Anthropometrics measurements · Skinfold thinkness · k-means clustering

1 Introduction

The increased prevalence of overweight and obesity has become a major factor in public spending in countries around the world [6]. Studies estimate that 57.8% of the world population will be overweight or obese by 2030 if current trends continue [6]. The obesity is commonly associated with several metabolic dysfunctions, such as insulin resistance [2,36], metabolic syndrome [29,35], increased blood glucose [1], dyslipidemia, hypertension and the development of other diseases such as type 2 diabetes, cardiovascular diseases [11,20] and atherosclerosis [1].

World Health Organization (WHO) has defined obesity as "an abnormal or excessive fat accumulation that presents a risk to health" [13]. Currently the diagnosis of overweight and obesity is based on body mass index (BMI). The WHO (2004) [37] proposed the cut-off points for defining underweight, normal weight, overweight and obesity in their different degrees. Among the limitations of BMI is the impossibility of discriminating between fatty tissue and muscle tissue, tending to produce false negatives in people with a high percentage of body fat but a normal BMI, and false positives in people with high BMI and high muscle tissue [3,23], because of this the use in concomitant of the body fat percentage (BFP) and BMI is recommended for a obesity diagnosis.

The BFP is calculated from several methods, among them are the bioelectric impedance and formula of Siri [33] that uses two, four and seven different skinfolds as variables, in these research the Siri formula with two skinfolds were used to compute the BFP [10]. Currently, there are no established limits for the abnormal BFP, mainly due to the limitation of the existing data around the world. Numerous studies have evaluated the relationship between overweight and obesity with BMI through skinfold thinness, finding a high directly proportional correlation between BMI, BFP and skinfold thinness [4,26].

Machine learning techniques to classify overweight and obesity have been already used [7]. Certain studies have used k-means to differentiate overweight and obesity from normal subjects using biochemical variables [21]. Some other studies use the k-means to detect overweight populations, based on anthropometric measures such as waist and hip circumference [9] and indicators of comorbidity such as diabetes, depression and atherosclerosis.

The aim of this work is to evaluate the k-means clustering algorithm using anthropometric measures to classify subjects with obesity and abnormal BFP. A database of 1053 subjects with anthropometric measurement (weight, height, arm circumferences, flexed arm circumferences, waist circumference, hip circumference, thigh circumferences, calf circumferences, triceps skinfolds, subscapular skinfolds, suprailiac skinfolds, abdominal skinfolds, thigh skinfolds, calf skinfolds, diameter of humerus and diameter of femur) values was used. In the following section the database and k-means method used in this investigation will be explained. In the Sects. 3 and 4 the results and discussion will be presented. And finally, in Sect. 5, conclusions and proposals for future work will be presented.

2 Methodology

2.1 Database

Between 2004 and 2012 [16], 1053 (male = 308) adult men and women from the district capital of Venezuela were recruited into the Nutritional Assessment Laboratory of the Simón Bolívar University. Anthropometric measurements such as: height, weight, height, arm circumferences, flexed arm cicumferences, waist circumference, hip circumference, thigh circumferences, calf circumferences, triceps skinfolds, subscapular skinfolds, suprailiac skinfolds, abdominal skinfolds, thigh skinfolds, calf skinfolds, humerus diameters and femur diameters were performed on each subject.

The diagnosis of overweight was made using the WHO guidelines which state that an overweight person has a BMI greater than or equal to 25. From the group of overweight subjects, 23 participants had a BMI greater than or equal to 30, indicating that they suffer from obesity [28]. Both overweight and obese subjects were placed in the same group for this study since we wanted to classify subjects with dysfunctional weight values.

Since there are no established limits for the abnormal BFP, the diagnosis of abnormal BFP were made according to [8,18,27] that established a cut off points of BFP $<$ 25% for men and BFP $<$ 30% for women as the limit of normality, above these limits are considered abnormal BFP.

All the procedures carried out in the study were in accordance with the ethical standards of the Bioethics Committee of the Simón Bolívar University and the 1964 Declaration of Helsinki and its subsequent amendments or comparable ethical standards. All subjects accepted the study by signing an informed consent form. Table 1 shows the characteristics of the dataset used, describing the values of each of the anthropometric variables by their mean and standard deviation of both normal and overweight subjects. While Table 2 shows the characteristics of normal and abnormal BFP subjects of the dataset used.

2.2 k-means Implemented

k-means [15] is a method that divide n observations into k clusters. In the k-means algorithm each observation is allocated to a cluster with the nearest centroid using a distance function, then, the centroids in each cluster are calculated again. This process is repeated until the centroids are the same between each step, and the final clusters are established.

In this study k-means were applied to each anthropometric measurement as separate variables (except height and weight because the BMI use them as

Table 1. Anthropometrics variables characteristics for obesity and overweight.

Anthropometrics variables	Normal weight Male = 246, n = 883	Overweight/obesity[c] Male = 62, n = 170
Age [years]	20.940 ± 2.773[b]	21.130 ± 2.919
Weight[a] [Kg]	56.125 ± 8.304	73.582 ± 10.174
Height [cm]	162.831 ± 8.521	163.274 ± 9.171
Right arm circumference[a] [cm]	25.762 ± 2.493	30.461 ± 2.373
Left arm circumference[a] [cm]	25.650 ± 2.425	30.432 ± 2.427
Right flexed arm circumference[a] [cm]	26.818 ± 2.717	31.381 ± 2.642
Left flexed arm circumference[a] [cm]	26.581 ± 3.109	31.119 ± 2.679
Waist circumference[a] [cm]	70.301 ± 7.095	84.824 ± 9.271
Hip circumference[a] [cm]	91.806 ± 5.471	103.539 ± 6.237
Right thigh circumference[a] [cm]	45.236 ± 2.370	49.594 ± 2.532
Left thigh circumference[a] [cm]	44.699 ± 2.565	49.093 ± 2.536
Right calf circumference[a] [cm]	33.736 ± 2.370	38.094 ± 2.532
Left calf circumference[a] [cm]	33.699 ± 2.565	38.093 ± 2.536
Right triceps skinfold[a] [mm]	13.593 ± 4.852	19.012 ± 5.634
Left triceps skinfold[a] [mm]	13.384 ± 4.788	18.861 ± 5.664
Right subscapular skinfold[a] [mm]	12.657 ± 3.991	20.512 ± 5.616
Left subscapular skinfold[a] [mm]	12.768 ± 4.005	20.658 ± 5.595
Right suprailiac skinfold[a] [mm]	11.837 ± 5.073	20.151 ± 6.451
Left suprailiac skinfold[a] [mm]	11.869 ± 5.102	20.216 ± 6.522
Right abdominal skinfold[a] [mm]	22.037 ± 5.073	30.351 ± 6.451
Left abdominal skinfold[a] [mm]	22.869 ± 5.102	31.216 ± 6.522
Right thigh skinfold[a] [mm]	19.899 ± 5.136	25.660 ± 6.019
Left thigh skinfold[a] [mm]	20.705 ± 5.219	26.282 ± 5.799
Right calf skinfold[a] [mm]	13.099 ± 5.136	18.860 ± 6.019
Left calf skinfold[a] [mm]	13.505 ± 5.219	19.082 ± 5.799
Right humerus diameter epicondylar[a] [cm]	5.973 ± 0.640	6.296 ± 0.634
Left humerus diameter epicondylar[a] [cm]	5.976 ± 0.635	6.301 ± 0.635
Right femur diameter epicondylar[a] [cm]	8.831 ± 0.661	9.588 ± 0.703
Left femur diameter epicondylar[a] [cm]	8.828 ± 0.660	9.595 ± 0.720
Body mass index [Kg/m^2][a]	21.106 ± 2.092	27.525 ± 2.349

[a]Statistically significant difference (p-value < 0.05) between control and overweight/obesity.
[b]Average and standard deviation.
[c]The database has 23 subjects with obesity.

variables in the case of obesity and overweight diagnosis; and triceps and subscapularis skinfolds in the case of abnormal BFP because Siri formula used them as variables); the number of groups was set to two (k = 2), to assess the ability of each variable to classify between obese/overweight and normal weight subjects, and between normal and abnormal BFP subjects. The Euclidean squared

Table 2. Anthropometrics variables characteristics for body fat percentage.

Anthropometrics variables	Normal BFP Male = 392, n = 949	Abnormal BFP Male = 16, n = 104
Age [years]	20.970 ± 2.790[b]	20.960 ± 2.840
Weight[a] [Kg]	57.770 ± 9.950	69.640 ± 11.950
Height [cm]	163.210 ± 8.610	160.130 ± 8.280
Right arm circumference[a] [cm]	26.161 ± 2.820	29.840 ± 2.760
Left arm circumference[a] [cm]	26.050 ± 2.770	29.840 ± 2.800
Right flexed arm circumference[a] [cm]	27.220 ± 3.040	30.570 ± 2.870
Left flexed arm circumference[a] [cm]	26.930 ± 3.000	30.360 ± 2.800
Waist circumference[a] [cm]	71.570 ± 8.410	82.440 ± 10.250
Hip circumference[a] [cm]	92.730 ± 6.290	102.550 ± 7.690
Right thigh circumference[a] [cm]	45.630 ± 2.680	48.770 ± 3.120
Left thigh circumference[a] [cm]	45.100 ± 2.850	48.230 ± 3.180
Right calf circumference[a] [cm]	34.130 ± 2.680	37.270 ± 3.120
Left calf circumference[a] [cm]	34.100 ± 2.850	37.230 ± 3.180
Right triceps skinfold[a] [mm]	13.550 ± 4.680	22.880 ± 3.740
Left triceps skinfold[a] [mm]	13.330 ± 4.610	22.800 ± 3.610
Right subscapular skinfold[a] [mm]	12.790 ± 3.870	24.250 ± 4.040
Left subscapular skinfold[a] [mm]	12.900 ± 3.880	24.420 ± 3.930
Right suprailiac skinfold[a] [mm]	12.110 ± 5.220	22.900 ± 5.240
Left suprailiac skinfold[a] [mm]	12.140 ± 5.250	23.010 ± 5.260
Right abdominal skinfold[a] [mm]	22.310 ± 5.220	33.100 ± 5.240
Left abdominal skinfold[a] [mm]	23.140 ± 5.250	34.010 ± 5.260
Right thigh skinfold[a] [mm]	19.990 ± 5.070	28.490 ± 5.400
Left thigh skinfold[a] [mm]	20.790 ± 5.130	29.050 ± 5.190
Right calf skinfold[a] [mm]	13.190 ± 5.070	21.690 ± 5.400
Left calf skinfold[a] [mm]	13.590 ± 5.130	21.850 ± 5.190
Right humerus diameter epicondylar [cm]	6.020 ± 0.650	6.090 ± 0.610
Left humerus diameter epicondylar [cm]	6.020 ± 0.650	6.100 ± 0.590
Right femur diameter epicondylar[a] [cm]	8.900 ± 0.700	9.400 ± 0.780
Left femur diameter epicondylar[a] [cm]	8.900 ± 0.700	9.410 ± 0.790
Body mass index [Kg/m^2][a]	21.600 ± 2.660	27.050 ± 3.360
Fat body percentage [%][a]	22.610 ± 6.090	33.000 ± 3.220

[a]Statistically significant difference (p-value < 0.05) between normal BFP and abnormal BFP.
[b]Average and standard deviation.

distance were used to calculate the distance between each variable of data set with centroids and the process were replayed 10 times to prevent local minima. The silhouette coefficient (SC) was used to assess the assignment of the data set in the respective cluster [32].

2.3 Metrics Calculation

The confusion matrix [12,31] is a table that contrasts the real classification with the classification made by the clustering model. In the Table 3 an example of confusion matrix is showed, the columns ($Class_1$, ..., $Class_n$) represent the k-means classification and the rows represent the real classification. The numbers in the main diagonal ($A_{11}, ..., A_{nn}$) are the right k-means method classification and n is the amount of the total classes. In this study, the objective is to classify obese subjects from normal weight subjects, and normal BFP subjects from abnormal BFP subjects, as a consequence of that, the number of classes is two ($n = 2$).

Table 3. Confusion Matrix.

True/predicted	$Class_1$	$Class_2$	$\ldots$	$Class_n$
$Class_1$	A_{11}	A_{12}	$\ldots$	A_{1n}
$Class_2$	A_{21}	A_{22}	$\ldots$	A_{2n}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$Class_n$	A_{n1}	A_{n2}	$\ldots$	A_{nn}

The accuracy (Acc) [31] represents the rate between the correctly classified instances and the total. Equation 1 shows the expression of accuracy, where A_{ij} are the instances for $i = 1, ..., n$ and $j = 1, ..., n$, and n is the number of total classes.

$$Acc = \frac{\sum_{i=1}^{n} A_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}} \tag{1}$$

The precision (P_i) [31] of a $Class_i$ (see Eq. 2) represents the rate between of correctly classified instances of the $Class_i$ (A_{ii}) (true positives) and the total classifications of the $Class_i$ (A_{ji}). In this study the precision reported is the class precision average.

$$P_i = \frac{A_{ii}}{\sum_{j=1}^{n} A_{ji}} \tag{2}$$

The recall (R_i) [12] of a $Class_i$ (see Eq. 3) is the rate between the $Class_i$ (A_{ii}) correctly classified instances and the total number of instances that have the $Class_i$ as the true label (A_{ij}). The recall reported in this study is the average of the entire class recall.

$$R_i = \frac{A_{ii}}{\sum_{j=1}^{n} A_{ij}} \tag{3}$$

2.4 Statistical Analysis

To determine the differences between groups of two, the Wilcoxon non-parametric paired pair statistical test was used and a p-value$\leq$ 5% was considered to be statistically significant [22].

3 Results

Table 1 reports the anthropometric measurements of the normal weight and overweight/obese subjects. The database consists of 1053 subjects, 83.86% belong to the normal weight subjects group and 16.14% are overweight/obese. The classification of overweight/obesity was made according to the WHO, all subjects with $BMI \geq 25$ were classified as overweight. The 13,5% of the subjects who belong to the overweight/obesity group have $BMI \geq 30$ indicating that endurance obesity. Table 2 reports the anthropometric measurements of the normal and abnormal BFP subjects. The classification of abnormal BFP group were made according to [18,27], that established as abnormal $BFP \geq 25$ in men and $BFP \geq 30$ in woman. The 9.88% of the subjects of the database presents an abnormal BFP and the 90.12% have a normal BFP; the subjects with abnormal BFP has a $BMI \geq 25$ indicating that they also belongs to the overweight/obesity group.

Table 4 and Table 5 show the confusion matrix of the variables with the best performance in the k-means non-supervising clustering for overweight/obesity and abnormal BFP classifications, respectively. In addition, the silhouette coefficient (SC), accuracy (Acc), precision (P) and recall (R) coefficient for overweight/obesity and abnormal BFP diagnosis was reported for $k = 2$ as it is shown in Table 6 and Table 7, respectively. Figure 1 shows the assignment of individuals to cumulus clusters for $k = 2$, using the anthropometric measurements. The character X represents the centroids of each cluster.

4 Discussion

Table 1 shows the descriptive and anthropometric measurements of the normal weight and the overweight/obese subjects. All parameters showed significant differences between the groups, except for age and height. All skinfolds showed higher values in overweight/obese subjects compared to normal weight subjects. On the other hand, Table 2 shows the descriptive and anthropometric measurements of the normal and abnormal BFP subjects. All parameters showed significant differences between the groups, except for age, height and epicondylar humerus diameter. All skinfolds showed higher values in abnormal BFP subjects compared to normal BFP subjects. All those facts are expected since obese and higher BFP subjects tend to have a thicker adipose panicle than normal weight and BFP subjects [17,34].

The k-means clustering method (Table 4) is capable of classifying obese subjects from normal weight subjects with the following anthropometric measures: right arm circumference, left arm circumference, right subscapular skinfold, left subscapular skinfold, waist circumference and hip circumference, with a $Acc \geq 0.78$, a $P \geq 0.78$, and a $R \geq 0.68$. On the other hand, k-means clustering demonstrated that is capable of classifying subjects with normal and abnormal

Table 4. Confusion matrix of k-means non-supervised classification of the variables with the best performance in the prediction of obesity and overweight

Anthropometrics measures	Confusion matrix		
	True/predicted	Normal weight $n = 883$	Overweight/obesity $n = 170$
Right arm circumference	Cluster 1	680[a] (77.01%)[b]	14[c] (8.24%)[d]
	Cluster 2	203 (22.99%)	156 (91.76%)
Left arm circumference	Cluster 1	681 (77.12%)	15 (8.82%)
	Cluster 2	202 (22.88%)	155 (91.18%)
Waist circumference	Cluster 1	677 (76.67%)	26 (15.29%)
	Cluster 2	206 (23.33%)	144 (84.71%)
Hip circumference	Cluster 1	659 (74.63%)	8 (4.71%)
	Cluster 2	224 (25.37%)	162 (95.29%)
Right subscapular skinfold	Cluster 1	698 (79.05%)	37 (21.76%)
	Cluster 2	185 (20.95%)	133 (78.24%)
Left subscapular skinfold	Cluster 1	695 (78.71%)	38 (22.35%)
	Cluster 2	188 (21.29%)	132 (77.65%)

[a]Number of subjects in the control group and classified in a respective cluster.
[b]Percentage of subjects in the control group and classified in a respective cluster.
[c]Number of subjects in the overweight/obesity group and classified in a respective cluster.
[d]Percentage of subjects in the overweight/obesity group and classified in a respective cluster.

BFP (Table 5) with the following anthropometric measures: right arm circumference, left arm circumference, waist circumference, hip circumference, suprailiac and abdominal skinfolds, with a $Acc \geq 0.73$, a $P \geq 0.73$, and a $R \geq 0.64$. Acceptable levels of accuracy and precision indicate that the method is capable of classifying subjects with the two pathologies. Slightly lower recall values indicate that the method is able to classify cases with the disease but gives a series of false negatives. It can also be seen that the silhouette coefficient (SC) is greater than 0.5 in all cases, indicating that all subjects were classified into a group for each of the parameters. In the parameters with the best Acc, P and R values, $SC \geq 0.55$ (Table 6 and Table 7).

Figure 1 shows that subjects with high skinfold values were located in cluster 2 (red) and subjects with lower skinfold values in cluster 1 (blue). Furthermore, the cluster 1 is where the highest percentage of normal weight and BFP subjects are found and the cluster 2 is where the highest percentage of overweight/obese and abnormal BFP subjects are found. This may be due to the fact that overweight subjects have a thicker adipose panicle and higher BFP than normal weight and BFP subjects [19,34]. The same fact is observed in the case of waist and hip circumference, where the method places the subjects with the largest hip and waist circumference in cluster 2, which is the group with the highest percentage of overweight/obese and abnormal BFP subjects. It should be noted

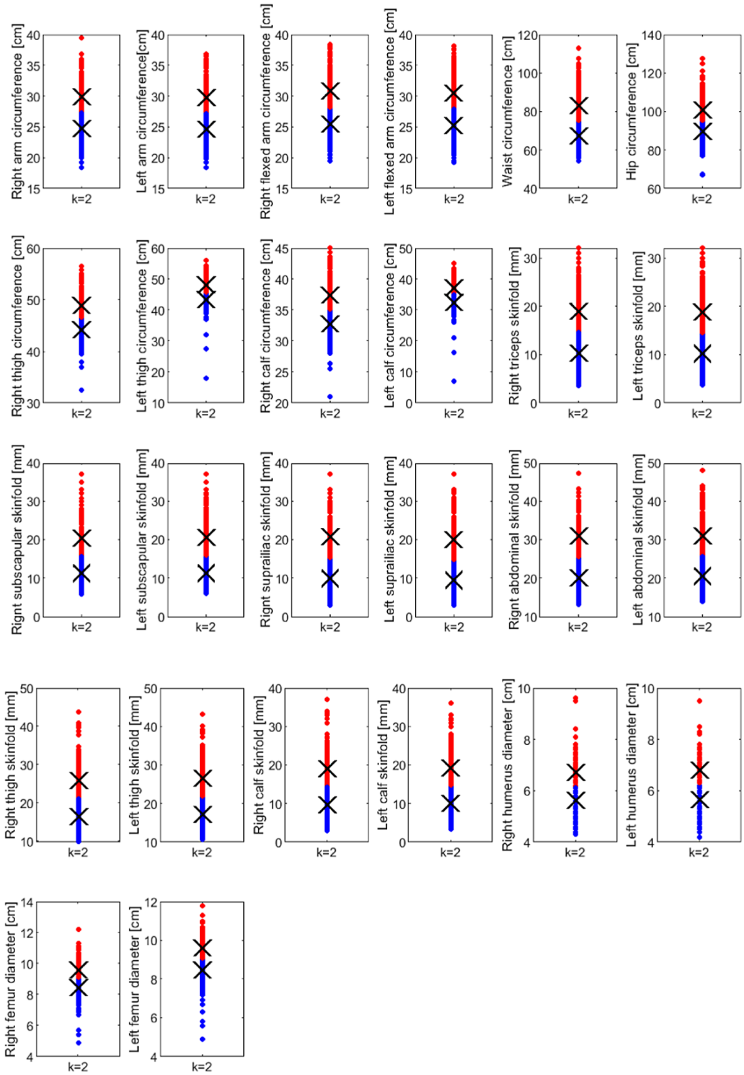

Fig. 1. Instance assignment (circles) to clusters for k = 2, using anthropometrics parameters. Red circles belong to cluster 1 and blue circles to cluster 2. Character X represents the cluster centroids. (Color figure online)

that waist circumference is strongly related to abdominal obesity and, in particular, it is used today as a risk factor for diseases such as cardiovascular disease and diabetes [5,39].

Table 5. Confusion matrix of k-means non-supervised classification of the variables with the best performance in the prediction of abnormal BFP

Anthropometrics measures	Confusion matrix		
	True/predicted	Normal BFP $n = 949$	Abnormal BFP $n = 104$
Right arm circumference	Cluster 1	673[a] (71.00%)[b]	21[c] (20.20%)[d]
	Cluster 2	276 (29.00%)	83 (79.80%)
Left arm circumference	Cluster 1	674 (71.00%)	22 (21.20%)
	Cluster 2	275 (29.00%)	82 (78.80%)
Waist circumference	Cluster 1	675 (71.10%)	28 (27.00%)
	Cluster 2	274 (28.90%)	76 (73.00%)
Hip circumference	Cluster 1	652 (68.70%)	15 (14.40%)
	Cluster 2	297 (31.30%)	89 (85.6%)
Right suprailiac skinfold	Cluster 1	717 (75.60%)	10 (9.60%)
	Cluster 2	232 (24.40%)	94 (90.40%)
Left suprailiac skinfold	Cluster 1	670 (70.60%)	7 (6.70%)
	Cluster 2	279 (29.40%)	97 (93.30%)
Right abdominal skinfold	Cluster 1	717 (75.60%)	10 (9.60%)
	Cluster 2	232 (24.40%)	94 (90.40%)
Left abdominal skinfold	Cluster 1	670 (70.60%)	7 (6.70%)
	Cluster 2	279 (29.40%)	97 (93.30%)

[a]Number of subjects in the normal BFP group and classified in a respective cluster.
[b]Percentage of subjects in the normal BFP group and classified in a respective cluster.
[c]Number of subjects in the abnormal BFP cluster.
[d]Percentage of subjects in the abnormal BFP group and classified in a respective cluster.

In the case of overweight/obesity diagnosis the arm circumference values show the best *Acc*, *P* and *R* (0.79, 0.84 and 0.71) compared to the all other measures. The subjects with the largest arm circumference were placed in cluster 2, which is the group with the highest percentage of overweight/obese subjects (91.76%). This result indicates a strong relationship between high arm circumferences and high BMI values, corroborating some studies [5,24]. On the other hand, in the case of abnormal BFP diagnosis, the right abdominal and suprailiac skinfolds show the best *Acc*, *P* and *R* (0.77, 0.68 and 0.64) compared to the all other measures. The subjects with largest suprailiac and abdominal skinfold were placed in the cluster 2 (right skinfold 90.40% and left skinfold 93.30%). This results are in concordance with studies that correlate high body fat percentage with high abdominal fat accumulation [25,30], especially in subjects with insulin resistance and high risk of develop type 2 diabetes [14,38].

Table 6. Silhouette coefficient (SC), accuracy (Acc), precision (P) and recall (R) from k-means algorithm for overweight/obesity classification.

Anthropometrics variables	Silhouette Coefficient	Metrics		
		Accuracy	Precision	Recall
Right arm circumference	0.60 ± 0.16	0.79	0.84	0.71
Left arm circumference	0.60 ± 0.17	0.79	0.84	0.71
Right flexed arm circumference	0.58 ± 0.18	0.75	0.82	0.68
Left flexed arm circumference	0.58 ± 0.19	0.77	0.81	0.68
Waist circumference	0.59 ± 0.19	0.78	0.81	0.69
Hip circumference	0.56 ± 0.18	0.78	0.84	0.70
Right thigh circumference	0.56 ± 0.17	0.77	0.81	0.69
Left thigh circumference	0.55 ± 0.18	0.73	0.80	0.67
Right calf circumference	0.56 ± 0.17	0.77	0.81	0.69
Left calf circumference	0.55 ± 0.18	0.73	0.80	0.67
Right triceps skinfold	0.56 ± 0.17	0.62	0.69	0.60
Left triceps skinfold	0.56 ± 0.18	0.62	0.68	0.60
Right subscapular skinfold	0.64 ± 0.19	0.79	0.79	0.68
Left subscapular skinfold	0.63 ± 0.18	0.79	0.78	0.68
Right suprailiac skinfold	0.60 ± 0.19	0.75	0.73	0.65
Left suprailiac skinfold	0.59 ± 0.19	0.73	0.74	0.64
Right abdominal skinfold	0.60 ± 0.19	0.75	0.73	0.65
Left abdominal skinfold	0.59 ± 0.19	0.73	0.74	0.64
Right thigh skinfold	0.58 ± 0.18	0.62	0.69	0.60
Left thigh skinfold	0.58 ± 0.18	0.62	0.70	0.61
Right calf skinfold	0.58 ± 0.18	0.62	0.69	0.60
Left calf skinfold	0.58 ± 0.18	0.62	0.70	0.61
Right humerus diameter epicondylar	0.62 ± 0.17	0.65	0.62	0.57
Left humerus diameter epicondylar	0.62 ± 0.17	0.68	0.60	0.56
Right femur diameter epicondylar	0.56 ± 0.17	0.67	0.71	0.62
Left femur diameter epicondylar	0.56 ± 0.18	0.66	0.71	0.61

Table 7. Silhouette coefficient (SC), accuracy (Acc), precision (P) and recall (R) from k-means algorithm for abnormal BFP classification.

Anthropometrics variables	Silhouette Coefficient	Metrics		
		Accuracy	Precision	Recall
Right arm circumference	0.60 ± 0.16	0.72	0.63	0.60
Left arm circumference	0.60 ± 0.17	0.72	0.62	0.60
Right flexed arm circumference	0.58 ± 0.18	0.67	0.60	0.59
Left flexed arm circumference	0.58 ± 0.19	0.67	0.60	0.59
Waist circumference	0.59 ± 0.19	0.71	0.61	0.59
Hip circumference	0.56 ± 0.18	0.70	0.63	0.60
Right thigh circumference	0.56 ± 0.17	0.69	0.59	0.58
Left thigh circumference	0.55 ± 0.18	0.64	0.57	0.57
Right calf circumference	0.56 ± 0.17	0.69	0.59	0.58
Left calf circumference	0.55 ± 0.18	0.64	0.57	0.57
Right suprailiac skinfold	0.60 ± 0.19	0.77	0.68	0.64
Left suprailiac skinfold	0.59 ± 0.19	0.73	0.66	0.62
Right abdominal skinfold	0.60 ± 0.19	0.77	0.68	0.64
Left abdominal skinfold	0.59 ± 0.19	0.73	0.66	0.62
Right thigh skinfold	0.58 ± 0.18	0.62	0.69	0.60
Left thigh skinfold	0.58 ± 0.18	0.61	0.59	0.59
Right calf skinfold	0.58 ± 0.18	0.61	0.59	0.59
Left calf skinfold	0.58 ± 0.18	0.61	0.59	0.59
Right humerus diameter epicondylar	0.62 ± 0.17	0.62	0.47	0.51
Left humerus diameter epicondylar	0.62 ± 0.17	0.65	0.48	0.51
Right femur diameter epicondylar	0.56 ± 0.17	0.61	0.53	0.55
Left femur diameter epicondylar	0.56 ± 0.18	0.61	0.52	0.55

5 Conclusions

The findings of this research suggest that the k-means method applied on anthropometric measurements can classify overweight/obese subjects and subjects with abnormal body fat percentage. The best anthropometric measurements to classify overweight and obesity on this research were: Arm circumferences, subscapular skinfolds, waist circumference and hip circumference. On the other hand, the best anthropometric measurements to classify abnormal BFP subjects in this research were: Arm circumferences, waist circumference, hip circumference, suprailiac and abdominal skinfolds.

Machine learning techniques, such as fully connected neural networks and the support vector machine, will be studied in the future to assess the relationship between BMI, BFP and anthropometric measurements. A machine learning technique would allow to evaluate how much influence have every antropometric variables over the classification, and would be possible to extract a spectrum to see which groups of subject are more vulnerable of suffering abnormal BFP.

Acknowledgment. This work was funded by the Research and Development Deanery of the Simón Bolívar University (DID) and the Research Direction of the Ibagué University. Full acknowledgement is given to David Powers, author of "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation" (BioInfo Publications™).

References

1. Ahamad, M.G., Ahmed, M.F., Uddin, M.Y.: Clustering as data mining technique in risk factors analysis of diabetes, hypertension and obesity. Eur. J. Eng. Res. Sci. **1**(6), 88–93 (2016)
2. Altuve, M., Severeyn, E., Wong, S.: Unsupervised subjects classification using insulin and glucose data for insulin resistance assessment. In: 2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA), pp. 1–7 (2015)
3. Altuve, M., Severeyn, E., Wong, S.: Adaptation of five indirect insulin sensitivity evaluation methods to three populations: metabolic syndrome, athletic and normal subjects. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4555–4558. IEEE (2014)
4. Bratke, H., et al.: Timing of menarche in Norwegian girls: associations with body mass index, waist circumference and skinfold thickness. BMC Pediatr. **17**(1), 138 (2017)
5. Chaput, J.P., et al.: Mid-upper arm circumference as a screening tool for identifying children with obesity: a 12-country study. Pediatr. Obes. **12**(6), 439–445 (2017)
6. Chooi, Y.C., Ding, C., Magkos, F.: The epidemiology of obesity. Metabolism **92**, 6–10 (2019)
7. DeGregory, K.W., et al.: A review of machine learning in obesity. Obes. Rev. **19**(5), 668–685 (2018)
8. Centers for Disease Control and Prevention: Quickstats: mean percentage body fat, by age group and sex–national health and nutrition examination survey, United States, 1999–2004 (2008)
9. Doménech-Asensi, G., Gómez-Gallego, C., Ros-Berruezo, G., García-Alonso, F.J., Canteras-Jordana, M.: Critical overview of current anthropometric methods in comparison with a new index to make early detection of overweight in Spanish university students: the normalized weight-adjusted index. Nutricion hospitalaria **35**(2), 359–390 (2018)
10. Durnin, J.V.G.A., Womersley, J.: Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. Br. J. Nutr. **32**(1), 77–97 (1974)
11. Farina, P.V.R., Severeyn, E., Wong, S., Turiel, J.P.: Study of cardiac repolarization during oral glucose tolerance test in metabolic syndrome patients. In: 2012 Computing in Cardiology, pp. 429–432. IEEE (2012)
12. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861–874 (2006)
13. Flegal, K.M., Carroll, M.D., Kit, B.K., Ogden, C.L.: Prevalence of obesity and trends in the distribution of body mass index among us adults, 1999–2010. Jama **307**(5), 491–497 (2012)

14. Genske, F., et al.: Abdominal fat deposits determined by magnetic resonance imaging in relation to leptin and vaspin levels as well as insulin resistance in the general adult population. Int. J. Obes. **42**(2), 183–189 (2018)
15. Hartigan, J.A., Wong, M.A.: Algorithm as 136: a k-means clustering algorithm. J. R. Stat. Soc. Ser. C (Appl. Stat.) **28**(1), 100–108 (1979)
16. Herrera, H., Rebato, E., Arechabaleta, G., Lagrange, H., Salces, I., Susanne, C.: Body mass index and energy intake in Venezuelan university students. Nutr. Res. **23**(3), 389–400 (2003)
17. Hung, S.P., Chen, C.Y., Guo, F.R., Chang, C.I., Jan, C.F.: Combine body mass index and body fat percentage measures to improve the accuracy of obesity screening in young adults. Obes. Res. Clin. Pract. **11**(1), 11–18 (2017)
18. Kim, J.Y., Han, S.H., Yang, B.M.: Implication of high-body-fat percentage on cardiometabolic risk in middle-aged, healthy, normal-weight adults. Obesity **21**(8), 1571–1577 (2013)
19. Krebs, N.F., Himes, J.H., Jacobson, D., Nicklas, T.A., Guilday, P., Styne, D.: Assessment of child and adolescent overweight and obesity. Pediatrics **120**(Supplement 4), S193–S228 (2007)
20. Ledezma, C.A., Perpiñan, G., Severeyn, E., Altuve, M.: Data fusion for QRS complex detection in multi-lead electrocardiogram recordings. In: 11th International Symposium on Medical Information Processing and Analysis, vol. 9681, p. 968118. International Society for Optics and Photonics (2015)
21. Li, L., Song, Q., Yang, X.: K-means clustering of overweight and obese population using quantile-transformed metabolic data. Diabetes Metab. Syndr. Obes. Targets Ther. **12**, 1573–1582 (2019)
22. Marusteri, M., Bacarea, V.: Comparing groups for statistical differences: how to choose the right statistical test? Biochemia Medica **20**(1), 15–32 (2010)
23. Mathew, H., Farr, O.M., Mantzoros, C.S.: Metabolic health and weight: understanding metabolically unhealthy normal weight or metabolically healthy obese patients. Metab. Clin. Exp. **65**(1), 73–80 (2016)
24. Mazicioglu, M.M., Hatipoglu, N., Öztürk, A., Cicek, B., Üstünbas, H.B., Kurtoglu, S.: Waist circumference and mid-upper arm circumference in evaluation of obesity in children aged between 6 and 17 years. J. Clin. Res. Pediatr. Endocrinol. **2**(4), 144 (2010)
25. Merrill, Z., Chambers, A., Cham, R.: Development and validation of body fat prediction models in American adults. Obes. Sci. Pract. **6**(2), 189–195 (2020)
26. Ojo, G., Adetola, O.: The relationship between skinfold thickness and body mass index in estimating body fat percentage on Bowen university students. Int. Biol. Biomed. J. **3**(3), 138–144 (2017)
27. Okorodudu, D., et al.: Diagnostic performance of body mass index to identify obesity as defined by body adiposity: a systematic review and meta-analysis. Int. J. Obes. **34**(5), 791–799 (2010)
28. World Health Organization: Physical inactivity: a global public health problem (2008)
29. Perpiñan, G., Severeyn, E., Altuve, M., Wong, S.: Classification of metabolic syndrome subjects and marathon runners with the k-means algorithm using heart rate variability features. In: 2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA), pp. 1–6. IEEE (2016)
30. Philipsen, A., et al.: Associations between ultrasound measures of abdominal fat distribution and indices of glucose metabolism in a population at high risk of type 2 diabetes: the addition-pro study. PloS One **10**(4), e0123062 (2015)

31. Powers, D.M.W.: Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation. J. Mach. Learn. Technol. **2**(1), 37–63 (2011)
32. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
33. Siri, W.E.: The gross composition of the body. In: Advances in Biological and Medical Physics, vol. 4, pp. 239–280. Elsevier (1956)
34. Suclla-Velásquez, J.A., Smedts, C.: Obesity: a risk factor for infection after surgery. In: Weight Management. IntechOpen (2010)
35. Velásquez, J., Herrera, H., Encalada, L., Wong, S., Severeyn, E.: Análisis dimensional de variables antropométricas y bioquímicas para diagnosticar el síndrome metabólico. Maskana **8**, 57–67 (2017)
36. Vintimilla, C., Wong, S., Astudillo-Salinas, F., Encalada, L., Severeyn, E.: An aide diagnosis system based on k-means for insulin resistance assessment in eldery people from the Ecuadorian highlands. In: 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), pp. 1–6. IEEE (2017)
37. Who, E.C.: Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet (Lond. Engl.) **363**(9403), 157 (2004)
38. Yang, H.R., Chang, E.J., et al.: Insulin resistance, body composition, and fat distribution in obese children with nonalcoholic fatty liver disease. Asia Pac. J. Clin. Nutr. **25**(1), 126 (2016)
39. Zhang, C., Rexrode, K.M., Dam, R.M.V., Li, T.Y., Hu, F.B.: Comparing groups for statistical differences: how to choose the right statistical test? Circulation **117**(13), 1658–1667 (2008)