Fuzzy Systems and Data Mining VII
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA210188

Support Vector Machine Technique as Classifier of Impaired Body Fat Percentage

Alexandra LA CRUZ [a,1], Erika SEVEREYN [b], Mónica HUERTA [c] and Sara WONG [d]

[a] *Faculty of Engineering, Universidad de Ibagué,Colombia*
[b] *Department of Thermodynamics and Transfer Phenomena, Universidad Simón Bolívar, Venezuela*
[c] *Faculty of Engineering, Universidad Politécnica Salesiana, Ecuador*
[d] *Department of Electronics and Circuits, Universidad Simón Bolívar, Venezuela*

Abstract. Excess weight and obesity are indicators of an unhealthy or harmful accumulation of fat that can be dangerous to health. Body mass index (BMI) refers to height-to-weight radio and is often used to identify overweight and obesity in adults. Although BMI is commonly used to diagnose obesity and overweight, it is ineffective in differentiating between high muscle mass and elevated body fat mass. Body fat percentage (BF%) is one of the best predictors of obesity because it quantifies adipose tissue. The Deurenberg equation is among the indirect methods to measure BF%; it uses BMI, age, and sex as parameters to calculate the BF%. Machine learning techniques demonstrated to be a good classifier of overweight, obesity, and diseases related to insulin resistance and metabolic syndrome. This study intends to evaluate anthropometric parameters as classifiers of BF% alteration using support vector machines and the Deurenberg equation for BF% estimation. The database used consisted of 1978 individuals with 24 different anthropometric measurements. The results suggest the SVM as a suitable technique for classifying individuals with normal and abnormal BF% values. Accuracy, F1 score, PPV, NPV, and sensitivity were above 0.8. Besides, the specificity value is below 0.7, which indicates that false positives may occur. As future work, this research intends to apply neural networks as a classification technique.

Keywords. Support vector machine, anthropometric measurements, fat body percentage, Monte Carlo cross validation

1. Introduction

Being overweight and obese are indicators of unhealthy or excessive fat accumulation on the body with the potential to be harmful to health [1]. Body mass index (BMI) is a simple measure of the height-to-weight ratio, commonly used to identify overweight and obesity among adults [2]. Although BMI is usually used to diagnose obesity and overweight, it has the disadvantage of not differentiating between high fat and lean mass [3]. Since those downsides with BMI, two classifications of obesity are reported in the literature,

[1] Corresponding Author: Faculty of Engineering Universidad de Ibagué, Carrera 22, Calle 67 B, Av. Ambalá, Ibagué, Tolima, Colombia; E-mail: alexandra.lacruz@unibague.edu.co.

the first classification is individuals with high body fat mass but the normal metabolic response, and the second classification endorse individuals that suffer from metabolic obesity with normal weight (MONW) [4,5]. The persons that are MONW have high cardiovascular and metabolic risk factors similar to individuals with abnormal BMI [6,7].

The body fat percentage (BF%) is the indicator that best predicts obesity due it can quantify tissue adipose [8]. Several direct methods measure BF%, including densitometry, dual-energy x-ray absorptiometry, bioelectrical impedance analysis, near-infrared reactance, dual-energy x-ray absorptiometry and magnetic resonance imaging. However, they are unsuitable for epidemiological studies since they are expensive, require specialized equipment and skilled professionals [9]. Some indirect methods to measure BF% use BMI, age, and sex as parameters to calculate the BF% applying the Deurenberg [10], Gallagher [11] and Jackson-Pollock [12] equations. There are high correlations between BF% and anthropometric measurements, among them the hip circumference, waist-to-height ratio, and the abdominal circumference [13]. Further, the SIRI equation calculates the BF% from the skin folds [14]. These studies have all determined an appropriate relationship between measurements of these anthropometric values and BF%.

There is no consensus regarding the cut-off points of BF% because there is no statistically representative database [15]. Some researches proposed 25% for men and 30% for women as a BF% cut off point [8,16]. Other researches suggest that age is a relevant factor to consider in the cut-off point establishment [17,18,19]. In this work, a spectral cut-off point was considered, including sex and age.

Machine learning techniques have been used to classify overweight, obesity, insulin resistance and metabolic syndrome [20,21]. Some studies have used support vector machines (SVM) and decision tree to differentiate individuals with and without metabolic syndrome from variables as waist circumference, waist to height ratio, body mass index, among others [20,22]. The k-means algorithm has also been used to detect individuals with insulin resistance and overweight using as variables waist and hip circumferences [21,23].

This study aims to assess the anthropometric variables as a classifier of impaired BF%. A database used consisted of 1978 individuals with 24 anthropometrics measures(weight, height, body circumferences, and body skinfolds). The SVM method evaluates the predictive ability of anthropometric measure variables. The next section describes the methodology. The results, and discussion are explained in sections 3 and 4. Finally, section 5 presents the conclusions and proposals for future work.

2. Methodology

2.1. Database

The Nutritional Evaluation Laboratory of the Simón Bolívar University collected during the period from 2004 to 2012 [24] the database used in this work. It counted with 1978 participants, of which 678 are men and the rest women. The implemented protocol performed 24 anthropometric measurements in each volunteer, including height, weight, body circumferences, and body folds. Additionally, the BF% is calculated using Deurenberg's equation. Deurenberg's equation, as can be seen in equation (1), uses BMI, age, and sex as variables. The sex variable is equal to one for men and equal to zero for women.

$$BF\% = 1.20\, BMI - 10.8\, sex - 0.23\, age - 5.4 \tag{1}$$

The clinical protocol implemented in the database followed the ethical standards of the Declaration of Helsinki of 1964 and the ethical standards of the ethical committee of the Simón Bolívar University. All participants accepted the conditions of the study by signing informed consent. Table 1 shows the average and standard deviation of anthropometric variables of the subjects with normal and altered BF%.

2.2. *Classifier Metrics*

For the detection of altered levels of BF%, 24 anthropometric measurements were used, weight and height were excluded since they are variables of the Deurenberg equation. The true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) were measured. It visualizes the classification discrepancies in the classifier model [25]. The accuracy (ACC), specificity (SPE), sensitivity (SEN), positive predictive value (PPV), negative predictive value (NPV), and F1 score (F1) were calculated using the equations (2), (3), (4), (5), (6), and (7) respectively.

$$ACC = \frac{(TP+TN)}{(TP+FP+TN+FN)} \tag{2}$$

$$SEN = \frac{TP}{(TP+FN)} \tag{3}$$

$$SPE = \frac{TN}{(TN+FP)} \tag{4}$$

$$PPV = \frac{TP}{(FP+TP)} \tag{5}$$

$$NPV = \frac{TN}{(FN+TN)} \tag{6}$$

$$F1 = 2\,\frac{(PPV)\,(SEN)}{(PPV+SEN)} \tag{7}$$

2.3. *Implementation of the support vector machine method*

The classification-regression method SVM is a classification method used for binary, multiple classifications, and classification-regression problems. SVM has proven to be considered among the best classifiers over a wide range of scenarios, making it one of the benchmarks in both statistical learning and machine learning fields [26].

Support Vector Machine is based on the Maximal Margin Classifier, which turns on the hyperplane concept. In this work, the SVM method allows classifying individuals with normal and abnormal BF% values. For this purpose, a Monte Carlo Cross-Validation (MCCV) [27], and a Gaussian kernel [28] were used. Figure 1 shows the procedure applied in this work. The database was randomly (with a uniform probability) divided 80% for training with SVM and the remaining 20% to test the trained SVM and calculate the metrics. The process was performed 100 times, and the metrics were calculated in each iteration and then averaged.

2.4. *Statistical tests*

For the statistical analysis, the Mann-Whitney U test was used, since it was assumed that the samples are not paired and have a different distribution than the normal, and a p-value of less than 5% was considered statistically significant [29]. The Tables 1 and 2 are presented as mean and standard deviation values (mean $\pm$ standard deviation).

Table 1. Anthropometric parameters for individuals with normal and abnormal BF%.

Anthropometric parameter	Normal BF% (n=1037)	Impaired BF% (n=941)
Age[years][a]	24.678±13.832	71.358±18.386
Weigth[Kg][a]	56.064±9.348	62.243±12.867
Heigth[cm][a]	162.278±8.699	156.614±9.802
Right arm circumference [cm][a]	25.820±2.817	27.931±4.023
Left arm circumference[cm][a]	25.694±2.802	27.749±4.066
Right flexed arm circumference [cm][a]	26.828±3.026	28.543±4.063
Left flexed arm circumference [cm][a]	26.576±3.339	28.245±4.093
Waist circumference [cm][a]	71.119±7.571	88.127±11.557
Hip circumference [cm][a]	91.763±6.339	95.235±10.105
Right thigh circumference [cm][a]	44.947±3.321	45.629±5.950
Left thigh circumference [cm][a]	44.431±3.403	45.447±5.873
Right calf circumference [cm][a]	33.582±2.938	33.198±4.053
Left calf circumference [cm][a]	33.549±3.070	33.089±4.063
Right triceps fold [mm][a]	13.656±5.186	14.953±6.747
Left triceps fold [mm]	13.471±5.129	14.872±6.636
Right subscapular fold [mm][a]	12.920±4.569	17.293±7.510
Left subscapular fold [mm][a]	13.031±4.585	17.472±7.426
Right suprailiac fold [mm][a]	12.148±5.424	18.229±7.725
Left suprailiac fold [mm][a]	12.198±5.456	18.230±7.663
Right abdominal fold [mm][a]	22.034±5.927	25.393±8.917
Left abdominal fold [mm][a]	22.805±6.033	25.318±9.048
Right thigh fold [mm][a]	19.770±5.847	20.887±9.452
Left thigh fold [mm][a]	20.530±5.917	21.047±9.532
Right calf fold [mm][a]	13.184±5.429	15.544±7.832
Left calf fold [mm][a]	13.591±5.468	15.891±7.816
BMI[Kg/m^2][a]	21.211±2.551	25.319±4.460
BF%[a]	22.667±5.901	36.988±8.609
[a]Statistically significant difference (p-value $<$ 0.05).		

Table 2. Metrics of the support vector machine classification and Monte Carlo Cross Validation.

Metrics	BF%
Sensibility	0.965±0.012
Specificity	0.679±0.044
Accuracy	0.897±0.015
F1 Score	0.935±0.010
NPV	0.855±0.043
PPV	0.907±0.016

3. Results

Table 1 presents the anthropometric values of control individuals and individuals with impaired BF%. The database includes 1978 individuals, 52.43% belongs to the control group, and 47.57% endures impaired BF%. The classification was performed according

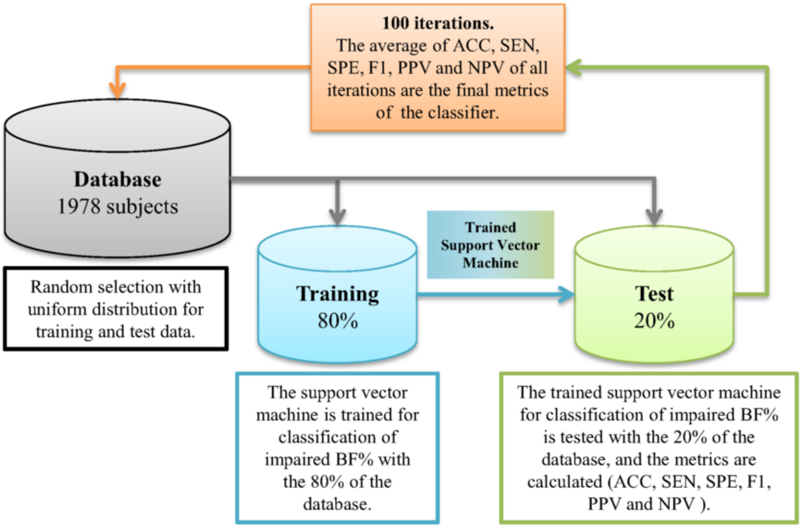

Figure 1. General methodology schematics for support vector machine classification.

to [11,18,17], whereby the cut-off point of the impaired BF% established for women between 20 and 39 years old is above 32%, for women between 40 and 59 years old 34%, and women over 60 years old 35%. On the other hand, the cut-off point of the impaired BF% in men between 20 and 39 years old is above 20%, men between 40 and 59 years old 22%, and men over 60 years old 23%. Table 2 reports the results of the training with the SVM using the 26 anthropometrics parameters displaying the average of sensitivity, specificity, accuracy, F1 score, negative, and positive predictive values performed 100 times.

4. Discussion

Table 1 shows features of subjects with normal and abnormal BF% values. Significant differences were found between groups in all anthropometric variables. Individuals with abnormal values of BF% have higher values of circumferences and folds than subjects with normal values of BF%. This finding corroborates studies that suggest that individuals with a high BF% have a thicker adipose pad than subjects with normal BF% [30,31]. Additionally, the average BMI of the individuals with abnormal BF% values suggests that they are overweight since it exceeds 25 Kg/m^2.

Furthermore, the waist circumference levels are normal for individuals with normal BF% values but are above 88 cm for individuals with abnormal BF% values. This finding could indicate that individuals with abnormal BF% values have a high accumulation of adipose tissue at the waist, which can be also observed in the abdominal fold, which is significantly higher in individuals with abnormal BF% than in individuals with normal BF% [32].

Table 2 shows the metrics of the SVM evaluation as a classifier of individuals with abnormal BF%. Almost all the metrics are above 0.8. The database used has approximately the same percentage of individuals with normal and abnormal BF%, suggesting that the accuracy is a valid metric to assess the capability of the methods to classify individuals with abnormal BF%. In this case, the capacity of the classifier method of classifying this metabolic dysfunction is high with an accuracy above 0.89 [33].

Moreover, the probabilities of obtaining a correct classification with individuals with normal and abnormal BF% values are high. Since the NPV, PPV, and F1 score values are above 0.85. The sensitivity value obtained above 0.96, indicates a false-negative rate of about 4%, which is low, and it is convenient for the classifier. Furthermore, specificity was the lowest metric value (< 0.7), indicating a high false-positive rate. Therefore, individuals with normal BF values may be classified as individuals with abnormal BF% values [34].

In other studies, automatic learning techniques have been used to detect other dysfunctions directly related to altered BF% levels. Farzaneh et. al. [20] used decision trees to determine alterations such as metabolic syndrome, obtaining a lower accuracy than the results obtained in this research (0.739 vs. 0.897), suggesting that SVM is the most suitable technique for detecting metabolic dysfunctions related to body composition than decision trees. Seyed-Taghi et. al. [35] used neural networks to classify obesity, obtaining sensitivity values lower than the sensitivity values found in this investigation (0.965 vs. 0.819). In contrast, the specificity values found in [35] are better than the values of specificity values found in this study (0.837 vs. 0.679), suggesting that the neural network technique should be used as future work.

5. Conclusion

In this research, a supervised machine learning technique (Support Vector Machine) was applied, and as a validation method, the Monte Carlo cross-validation technique was used. The results indicate that SVM was a reliable technique for classifying individuals based on body fat percentage (BF%), with an accuracy, F1 score, PPV, NPV, and sensitivity of more than 0.8.

Notwithstanding, the specificity value is less than 0.7, indicating that false positives may occur, this does not affect the classifier, considering that false negatives are the events to avoid. Further work will include the application of neural networks as a classification technique.

6. Acknowledgment

The Research and Development Deanery of the Universidad Politecnica Salesiana from Cuenca, Ecuador, the Research and Development Deanery (DID) of the Universidad Simón Bolívar, Caracas, Venezuela, and the Research Direction of Universidad de Ibagué from Ibagué Colombia are the main founded of this project.

References

[1] R. S. Chan and J. Woo, “Prevention of overweight and obesity: how effective is the current public health approach,” *International journal of environmental research and public health*, vol. 7, no. 3, pp. 765–783, 2010.

[2] C. B. Weir and A. Jan, “Bmi classification percentile and cut off points,” January 2019.

[3] J. Gómez-Ambrosi, C. Silva, J. Galofré, J. Escalada, S. Santos, D. Millán, N. Vila, P. Ibañez, M. Gil, V. Valentí, *et al.*, “Body mass index classification misses subjects with increased cardiometabolic risk factors related to elevated adiposity,” *International journal of obesity*, vol. 36, no. 2, pp. 286–294, 2012.

[4] E. A. Sims, "Are there persons who are obese, but metabolically healthy?," *Metabolism-Clinical and Experimental*, vol. 50, no. 12, pp. 1499–1504, 2001.

[5] S. Gómez-Zorita, M. Queralt, M. A. Vicente, M. González, and M. P. Portillo, "Metabolically healthy obesity and metabolically obese normal weight: a review," *Journal of physiology and biochemistry*, pp. 1–15, 2021.

[6] M. Garg and S. Mohale, "Prevalence of metabolic obesity normal weight (monw) in cardiovascular disease patients–a hospital-based case control study.," *Journal of Evolution of Medical and Dental Sciences*, vol. 9, no. 34, pp. 2427–2432, 2020.

[7] H. S. Sachdev, A. Porwal, A. Sarna, R. Acharya, S. Ramesh, U. Kapil, and A. V. Kurpad, "Intraindividual double-burden of anthropometric undernutrition and "metabolic obesity" in indian children: a paradox that needs action," *European Journal of Clinical Nutrition*, pp. 1–13, 2021.

[8] P. Macek, M. Biskup, M. Terek-Derszniak, M. Stachura, H. Krol, S. Gozdz, and M. Zak, "Optimal body fat percentage cut-off values in predicting the obesity-related cardiovascular risk factors: A cross-sectional cohort study," *Diabetes, metabolic syndrome and obesity: targets and therapy*, vol. 13, p. 1587, 2020.

[9] K. González-Ruíz, M. Medrano, J. E. Correa-Bautista, A. García-Hermoso, D. H. Prieto-Benavides, A. Tordecilla-Sanders, C. Agostinis-Sobrinho, M. Correa-Rodríguez, J. Schmidt Rio-Valle, E. González-Jiménez, *et al.*, "Comparison of bioelectrical impedance analysis, slaughter skinfold-thickness equations, and dual-energy x-ray absorptiometry for estimating body fat percentage in colombian children and adolescents with excess of adiposity," *Nutrients*, vol. 10, no. 8, p. 1086, 2018.

[10] P. Deurenberg, A. Andreoli, P. Borg, K. Kukkonen-Harjula, A. De Lorenzo, W. D. Van Marken Lichtenbelt, G. Testolin, R. Vigano, and N. Vollaard, "The validity of predicted body fat percentage from body mass index and from impedance in samples of five european populations," *European journal of clinical nutrition*, vol. 55, no. 11, pp. 973–979, 2001.

[11] D. Gallagher, M. Visser, D. Sepulveda, R. N. Pierson, T. Harris, and S. B. Heymsfield, "How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?," *American journal of epidemiology*, vol. 143, no. 3, pp. 228–239, 1996.

[12] A. S. Jackson, M. L. Pollock, and L. R. Gettman, "Intertester reliability of selected skinfold and circumference measurements and percent fat estimates," *Research Quarterly. American Alliance for Health, Physical Education and Recreation*, vol. 49, no. 4, pp. 546–551, 1978.

[13] F. Amirabdollahian and F. Haghighatdoost, "Anthropometric indicators of adiposity related to body weight and body shape as cardiometabolic risk predictors in british young adults: superiority of waist-to-height ratio," *Journal of obesity*, vol. 2018, 2018.

[14] M. Arroyo, A. M. Rocandio, L. Ansotegui, H. Herrera, I. Salces, and E. Rebato, "Comparison of predicted body fat percentage from anthropometric methods and from impedance in university students," *British Journal of Nutrition*, vol. 92, no. 5, pp. 827–832, 2004.

[15] B. Dong, Y. Peng, Z. Wang, O. Adegbija, J. Hu, J. Ma, and Y.-H. Ma, "Joint association between body fat and its distribution with all-cause mortality: A data linkage cohort study based on nhanes (1988-2011)," *PLoS One*, vol. 13, no. 2, p. e0193368, 2018.

[16] A. Jia, S. Xu, J. Ming, Y. Xing, J. Guo, M. Zhao, L. Zhang, and Q. Ji, "Body fat percentage cutoffs for risk of cardiometabolic abnormalities in the chinese adult population: a nationwide study," *European journal of clinical nutrition*, vol. 72, no. 5, pp. 728–735, 2018.

[17] H. Gharote, K. Pushpanshu, R. Kaushik, and R. Gharote, "Estimation of body fat percentage in oral squamous cell carcinoma," *International healthcare research journal*, vol. 2-11, no. 1, pp. 291–296, 2019.

[18] P. Deurenberg, J. A. Weststrate, and J. C. Seidell, "Body mass index as a measure of body fatness: age-and sex-specific prediction formulas," *British journal of nutrition*, vol. 65, no. 2, pp. 105–114, 1991.

[19] D. Gallagher, S. B. Heymsfield, M. Heo, S. A. Jebb, P. R. Murgatroyd, and Y. Sakamoto, "Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index," *The American journal of clinical nutrition*, vol. 72, no. 3, pp. 694–701, 2000.

[20] F. Karimi-Alavijeh, S. Jalili, and M. Sadeghi, "Predicting metabolic syndrome using decision tree and support vector machine methods," *ARYA atherosclerosis*, vol. 12, no. 3, p. 146, 2016.

[21] J. Velásquez, S. Wong, L. Encalada, H. Herrera, and E. Severeyn, "Lipid-anthropometric index optimization for insulin sensitivity estimation," in *11th international symposium on medical information processing and analysis*, vol. 9681, p. 96810R, International Society for Optics and Photonics, 2015.

[22] A. Worachartcheewan, N. Schaduangrat, V. Prachayasittikul, and C. Nantasenamat, "Data mining for

the identification of metabolic syndrome status," *EXCLI journal*, vol. 17, p. 72, 2018.
[23] C. Vintimilla, S. Wong, F. Astudillo-Salinas, L. Encalada, and E. Severeyn, "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.
[24] H. Herrera, E. Rebato, G. Arechabaleta, H. Lagrange, I. Salces, and C. Susanne, "Body mass index and energy intake in venezuelan university students," *Nutrition research*, vol. 23, no. 3, pp. 389–400, 2003.
[25] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Information sciences*, vol. 507, pp. 772–794, 2020.
[26] W. S. Noble, "What is a support vector machine?," *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
[27] Q.-S. Xu and Y.-Z. Liang, "Monte carlo cross validation," *Chemometrics and Intelligent Laboratory Systems*, vol. 56, no. 1, pp. 1–11, 2001.
[28] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural computation*, vol. 15, no. 7, pp. 1667–1689, 2003.
[29] M. Marusteri and V. Bacarea, "Comparing groups for statistical differences: how to choose the right statistical test?," *Biochemia medica: Biochemia medica*, vol. 20, no. 1, pp. 15–32, 2010.
[30] W. L. Ripka, L. Ulbricht, and P. M. Gewehr, "Body composition and prediction equations using skinfold thickness for body fat percentage in southern brazilian adolescents," *Plos one*, vol. 12, no. 9, p. e0184854, 2017.
[31] G. Ojo and O. Adetola, "The relationship between skinfold thickness and body mass index in estimating body fat percentage on bowen university students," *International Biological and Biomedical Journal*, vol. 3, no. 3, pp. 138–144, 2017.
[32] K. M. Flegal, J. A. Shepherd, A. C. Looker, B. I. Graubard, L. G. Borrud, C. L. Ogden, T. B. Harris, J. E. Everhart, and N. Schenker, "Comparisons of percentage body fat, body mass index, waist circumference, and waist-stature ratio in adults," *The American journal of clinical nutrition*, vol. 89, no. 2, pp. 500–508, 2009.
[33] K. Michael, M. P. Garcia-Souto, and P. Dabnichki, "An investigation of the suitability of artificial neural networks for the prediction of core and local skin temperatures when trained with a large and gender-balanced database," *Applied Soft Computing*, vol. 50, pp. 327–343, 2017.
[34] A. G. Lalkhen and A. McCluskey, "Clinical tests: sensitivity and specificity," *Continuing education in anaesthesia critical care & pain*, vol. 8, no. 6, pp. 221–223, 2008.
[35] S. T. Heydari, S. M. T. Ayatollahi, and N. Zare, "Comparison of artificial neural networks with logistic regression for detection of obesity," *Journal of medical systems*, vol. 36, no. 4, pp. 2449–2454, 2012.