Classifying people based on fat by a Neuro-Fuzzy System

Mohammadreza Valizadeh[1*], Ali Karamshahi[2], Kurosh Djafarian[3], Akbar Azizifar[4]

1. Department of Computer and Information Technology, Ilam University, Ilam, Iran
2. Department of Computer and Information Technology, Islamic Azad University, Ilam, Iran
3. Department of Clinical nutrition, School of Nutritional Sciences and Dietetic, Tehran University of Medical Sciennces, Tehran, Iran
4. Department of English Language, School of Medicine, Medical University of Ilam, Ilam, Iran

***Corresponding author:** Tel: +98 9183416055, Fax: +98
Address: Department of Computer and Information Technology, Ilam University, 69315-516 Ilam, Iran
E-mail: mr.valizadeh@ilam.ac.ir
Received: 14/02/2020 Revised: 18/03/2020 Accepted: 13/05/2020

Abstract

Introduction: Using BIA for body fat calculation is a normal method. The body fat factor is one of the most useful measures for assessing the risk of obesity. In this research, people are classified based on body fat. This research does not use any device. Adaptive Network-based Fuzzy Inference System (ANFIS) which is widely used in medical sciences, has been used to predict the exact category of fat.
Materials and Methods: A nutrition clinic in Tehran has collected 610 samples from its patients. Each data has six attributes: age, height, weight, BMI, gender, and fat percentage. Based on percentage fat, people are divided into six fat classes from very low fat to very high fat. This research uses ANFIS system to estimate body fat class. Age, height, weight, BMI, and gender are used as inputs of the system and fat class as output. Furthermore, for evaluating the proposed method, precision method is used.
Results: This research used machine learning techniques (i.e., ANFIS) to predict the class of fat people without using costly tools. The data showed that our method has an accuracy of 90.83%.
Conclusion: The results of this research show that using ANFIS can estimate accurately the category of body fat without any device. Therefore, it reduces diagnosis price.

Keywords: Learning algorithm, Body fat category, Data mining, ANFIS

Introduction

Nowadays, one of the human beings' problems is obesity. In the world, obesity has been doubled from 1980 to 2013 (1). Several factors can increase obesity including: Easy access to high-energy foods and low physical activity. Obesity directly or indirectly can lead to a variety of illnesses such as hypertension, heart disease, diabetes, and stroke. The body fat factor is one of the most useful measures for assessing the risk of obesity (2). Nowadays, medical centers use medical equipment, such as Bioelectrical Impedance Analysis (BIA), to calculate body fat levels which costs high. For this reason, this article uses machine learning techniques to provide an efficient model for estimating people's body fat so that it has the least necessary accuracy and does not use high-cost tools as well. For this purpose, data mining techniques have been used (3, 4). They combine and use different sciences to extract hidden knowledge in the data. The neuro fuzzy system is one of these techniques.

The neuro fuzzy system is among the computational methods in machine learning for knowledge display and applying the knowledge to predict output of complex systems. The main idea of neuro fuzzy system is to combine fuzzy system and artificial neural network to process data for learning and creating knowledge.

One of the most well-known neuro fuzzy systems is Adaptive Network-based Fuzzy Inference System (ANFIS); a structure that consists of multiple layers of fuzzification layer, rule layer, normalization layer, de-fuzzification layer, and output layer. Each layer is directly connected to the next one. The appropriate structure of the system is selected based on input, membership degree, rules, input, and output membership functions. In the training phase, by modifying the membership degree parameters based on the acceptable error rate, the input values approach the actual values. One of the learning algorithms of ANFIS is neural networks (5).

ANFIS has been widely used in medical sciences. In (5, 6) ANFIS was used to remove various noises from medical images. Researchers in (7, 8) used ANFIS to segment medical images and diagnose breast cancer and brain tumors. Also, in 2011, Soni et al. (9) used neural networks to predict heart disease. They used a set of datasets prepared by Rajkumar (10). In 11 different methods, including neural network, have been used to predict breast cancer. They used cross validation method to evaluate the performance of their system. Kara et al. (12) used an artificial neural network to analyze the tests for diagnosing the optic nerve disease. The final results were classified as healthy and patient. Kumar et al. used random forests and artificial neural networks to diagnose kidney stones (13).

In this study, ANFIS was applied to classify people based on fat. This was made possible by data collected from people who went to the nutrition clinic. The results show the efficiency of the proposed method.

Materials and Methods

In this study, the data collected by one of the nutrition clinics in Tehran in 1993 and 1994 was used. The database contains 610 records. A neuro fuzzy network called ANFIS is used to predict which type of obesity a person belongs to. In this network, the number of input layer neurons is determined according to the number of features of each sample which is equal to five. Furthermore, output layer has a neuron. Due to the complexity of the image with 5 inputs, the structure of a network with two inputs is shown in Figure (1).

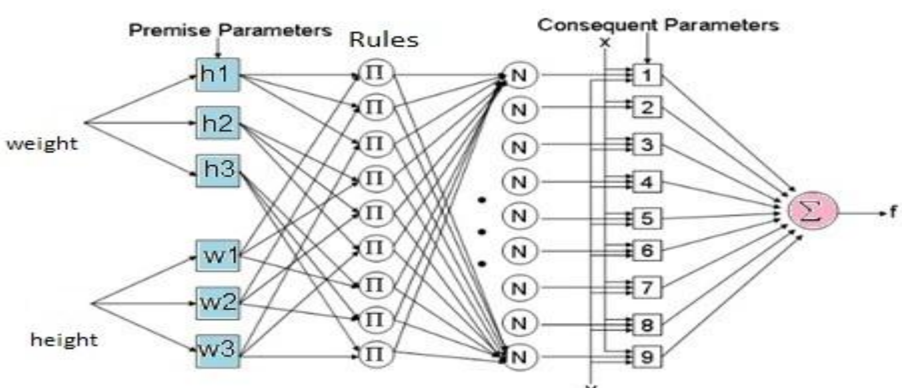

Figure 1. Structure of ANFIS with two of five inputs that is used in this study to determine the body fat class.

In Figure 1, For better understanding, only the two height and weight inputs have been drawn. In this system, weight is divided into three classes of light (w1), medium (w2), heavy (w3), and height in three classes of short (h1), medium (h2), and long (h3). Given the number of inputs, which is 2, and each has three subsets, then the next layer consists of nine rules. Then in the next layer, the rules are normalized. After that, in the next layer, the result of the rules is calculated based on height and weight (x and y are height and weight), and finally in the last layer, the results of the rules are integrated and the total output of the system is generated.

Before training, data were divided into very small, excellent, good, medium, high, and very high categories based on gender and body fat, as shown in Table 1.

Table 1. Classification of fat percentage to six fat classes for females. Each range of age has a different definition for six classes

Age class	Very high	High	Middle	Good	Excellent	Very low
Age<=19	Fat>32.1	27.1<=Fat<=32	22.1<=Fat<=27	17.1<=Fat<=22	12<=Fat<=17	Fat<12
20<=Age<=29	Fat>33.1	28.1<=Fat<=33	23.1<=Fat<=28	18.1<=Fat<=23	12<=Fat<=18	Fat<12
30<=Age<=39	Fat>34.1	29.1<=Fat<=34	24.1<=Fat<=29	19.1<=Fat<=24	12<=Fat<=19	Fat<12
40<=Age<=49	Fat>35.1	30.1<=Fat<=35	25.1<=Fat<=30	20.1<=Fat<=25	12<=Fat<=20	Fat<12
Age>=50	Fat>36.1	31.1<=Fat<=36	26.1<=Fat<=31	21.1<=Fat<=26	12<=Fat<=21	Fat<12

In the training phase, the ANFIS network structure is created and its parameters, which include the number of membership functions and their ranges, are optimally adjusted, for which the back propagation method is used. The created network is then evaluated. In order to accurately evaluate the proposed method, the accuracy measure has been used, which calculates the percentage of correct answers in the testing phase. Another important parameter that can be used in cross validation is Mean Square Error (MSE) calculated by Equation (1).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2 \quad (1)$$

where N is equal to the number of tested samples (test data is considered in ten rounds of testing), y_i is equal to the estimated fat category number for the sample number i by the proposed method and $\widehat{y_i}$ is the real fat category number for the mentioned sample. Also, measure of the Mean Absolute Error is calculated by the proposed method and determined by Equation (2).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y_i}| \quad (2)$$

where variables are mentioned for equation 1.

It is noteworthy that the proposed model has been implemented and tested in the MATLAB software environment.

Results

The dataset used in this study has 610 records. General information about the dataset is shown in Table 2. In the table the names of the features (i.e., attributes) in the dataset, maximum and minimum values in the dataset for each feature, standard deviation, and the average of the feature values are depicted. For example, the first row shows that the average of the age for 610 people is 35.4 and the maximum age is 77. BMI also represents the body mass index and is calculated by dividing the body weight by square of the height of each person. Furthermore, in the gender feature, the female gender is 0 and the male gender is 1. The goal of the proposed method is to achieve the least difference between the actual body fat category and the value estimated by the proposed model. For this purpose, MSE and MAE have been used to evaluate the efficiency of the proposed model.

Table 2. Dataset specifications used in this research. Each row shows statistical specifications for each one of the attributes.

Feature	Standard deviation	Average	Minimum	Maximum
Age (year)	10.9	35.4	15	77
Height(cm)	8.1	162.4	143	189
Weight(kg)	17.5	80.5	39.5	163.4
BMI	6.2	30.5	14.7	50.1
Gender	-	-	-	-
Fat percentage	9.7	35.2	2	52.1

As mentioned, the dataset has 610 data that 490 data were used for training and 120 for testing.

The goal of the proposed method is to achieve the least difference between the actual body fat category and the value estimated by the proposed model. For this purpose, MSE and MAE have been used to evaluate the efficiency of the proposed model. As mentioned, the dataset has 610 data that 490 data were used for training and 120 for testing. Figure 2 shows the data on a person's body fat group and the value estimated by the network for the test set. In the figure (Figure 2), the actual values are displayed in blue circles and the ANFIS output values are displayed in red. As the results of this experiment show, these points are matched, and the number of errors in the proposed system in estimating the exact body fat classes is low. To calculate the error of the proposed method, values of the actual class of samples are deducted from the values of estimated classes by ANFIS. The errors that are calculated for test data are shown in Figure 3. It shows that the average error of the system is 0.1 units for each test sample which is acceptable. Therefore, the proposed model has an accuracy of 90.83%. The experiment was done 10 times (i.e., system test was done 10 times with selecting training set and test set randomly) and MSE and MAE are 0.1166 and 0.1 respectively.

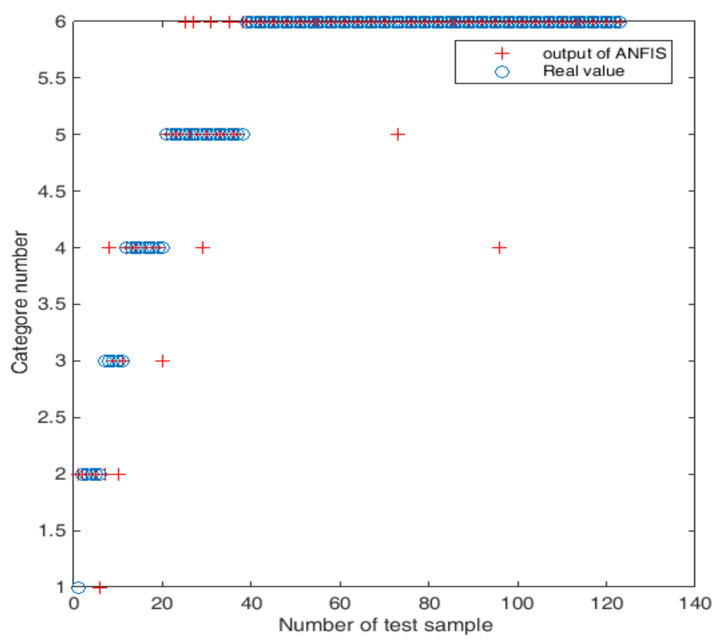

Figure 2. Estimated value for fat class in red and real value in blue for 120 test data.

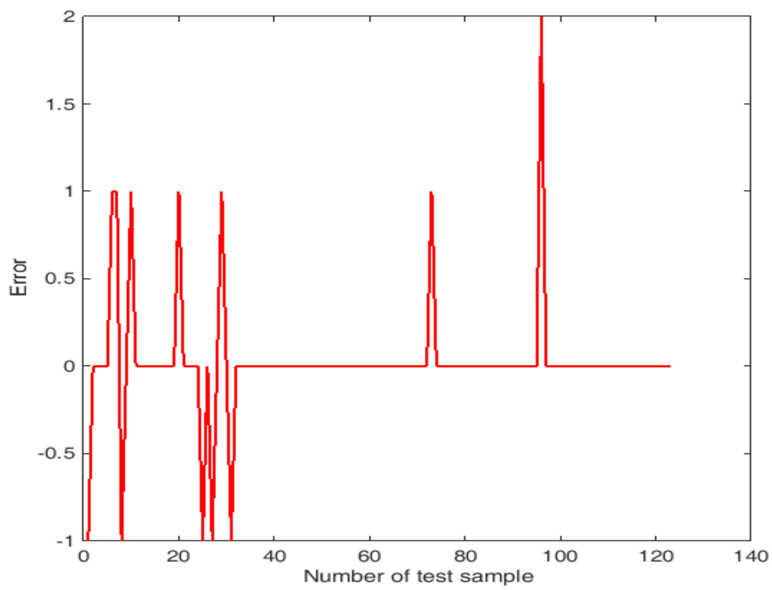

Figure 3. Amount of error for each test data. Data No. 100 has 2-unit error.

Discussion

Calculating the fat class of the human body is very important for which the doctors are trying special equipment. Smart tools that use machine learning techniques to predict the class of fat people without using costly tools can play an important role in people's health and therefore play an effective role in preventing chronic diseases.

Few studies have been done to predict the body fat category of humans using artificial intelligence and machine learning tools. In 2013, Kupusinac et al. used an artificial neural network to predict the percentage of fat in the human body. Their working accuracy is 80.43% (10). Compared to their method, the proposed method in this study is ANFIS. Although their samples are more (about 2,700), due to the greater number of features in the present study and the nature of the ANFIS system, their model is less accurate than the proposed method in this study.

The other research was done by Kehoe et al. who used bio-impedance to predict Indian students' fat percentage. It requires tools and spending money, which is unintelligent method. In contrast, the method proposed in this study does not require special tools and is cheap (15). Another study was conducted in 2016 on the fat of students in Mali based on skin thickness and analysis of bioelectrical impedance (16). The work has been done on 160 students and its results were remarkable, but the method is costly and time consuming. The other study was done by Raymond Chiong et al in 2020. It proposed an improved relative error support vector machine method to predict body fat. They believe that their method is cost-effective (17) and finally Robinson Ramírez-Vélez et al (18) investigated body fat percentage and fat mass index to prediction of metabolic syndrome among Colombian University students.

In this study, to accurately estimate a person's body fat category, a neuro fuzzy system was introduced. The proposed method average error is 0.1 unit per sample. The method only necessity is to obtain age, height, weight, gender, and BMI from the patient so that it can estimate his/her fat category. The results of the experiments show that the proposed model has the necessary accuracy to accurately estimate the body fat class of the human body and can be used as an efficient tool.

Conclusion

ANFIS is one of the well-known Machine Learning Techniques which are used widely. In this research ANFIS is used to classify fat of people based on their data without any specific tools. The results are promising and the proposed method accuracy is 90.83%. For future work, it is recommended to use clustering algorithms to estimate body fat. To sum up, the use of machine learning techniques in medical science can lead to interesting and significant results.

Acknowledgments

We appreciate all the colleagues and researchers who helped us to do this study.

References

1. Myint PK, Kwok CS, Luben RN, Wareham NJ, Khaw KT. Body fat percentage, body mass index and waist-to-hip ratio as predictors of mortality and cardiovascular disease. Heart. 2014;100(20):1613-9. doi: 10.1136/heartjnl-2014-305816.
2. Mohammadi S, Shakerhosseini R, Rastmanesh R, Jafarian K, Amiri Z, Jahangir F. Effects of melatonin supplementation on weight and body fat mass percentage in overweight or obese people. J Inflam Dis. 2015; 19 (5) :31-24.
3. Hand D, Mannila H, Smyth P. Principles of Data Mining. MIT Press. 2001; 128-15.
4. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques, 3rded. Elsevier Press. 2012; 368- 375.
5. A. Choubey, G. R. Sinha and S. Choubey, "A hybrid filtering technique in medical image denoising: Blending of neural network and fuzzy inference," 2011 3rd International Conference on Electronics Computer Technology, Kanyakumari, India, 2011, pp. 170-177, doi: 10.1109/ICECTECH.2011. 5941584.
6. Yüksel ME. A hybrid neuro-fuzzy filter for edge preserving restoration of images corrupted by impulse noise. IEEE Trans Image Process. 2006;15(4):928-36. doi: 10.1109/tip.2005.863941. PMID: 16579379.
7. W. Xu, L. Li and P. Xu, "A New ANN-based Detection Algorithm of the Masses in Digital Mammograms," 2007 IEEE International Conference on Integration Technology, Shenzhen, China, 2007, pp. 26-30, doi: 10.1109/ICITECHNOLOGY.2007.4290471.
8. Weidong XU, Wei LI, Lihua LI, Li MA, Shunren XI, Juan Z, et al. A New Computer-aided Detection System of CR Mammograms. J Comput Inf Syst. 2010; 6:2885-900.
9. Soni J, Ansari U, Sharma D, Soni S. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. IJCSE. 2011; 17:43-48.
10. Rajkumar A, Reena GS. Diagnosis of Heart Disease Using Data Mining Algorithm. GJCST. 2010; 10:38-43.
11. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med. 2005;34(2):113-27. doi: 10.1016/j.artmed.2004.07.002.
12. Kara S, Güven A, Oner AO. Utilization of artificial neural networks in the diagnosis of optic nerve diseases. Comput Biol Med. 2006;36(4):428-37. doi: 10.1016/j.compbiomed.2005.01.003.
13. Kumar M. Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm. Int J Comp Sci Mob Com. 2016(16); 5:24-33.
14. Kupusinac A, Stokić E, Doroslovački R. Predicting body fat percentage based on gender, age and BMI by using artificial neural networks. Comput Methods Programs Biomed. 2014;113(2):610-9.

doi: 10.1016/j.cmpb.2013.10.013. Epub 2013 Oct 29. PMID: 24275480.
15. Kehoe SH, Krishnaveni GV, Lubree HG, Wills AK, Guntupalli AM, Veena SR, Bhat DS, Kishore R, Fall CH, Yajnik CS, Kurpad A. Prediction of body-fat percentage from skinfold and bio-impedance measurements in Indian school children. Eur J Clin Nutr. 2011;65(12):1263-70. doi: 10.1038/ejcn.2011.119.
16. Noradilah MJ, Ang YN, Kamaruddin NA, Deurenberg P, Ismail MN, Poh BK. Assessing Body Fat of Children by Skinfold Thickness, Bioelectrical Impedance Analysis, and Dual-Energy X-Ray Absorptiometry: A Validation Study Among Malay Children Aged 7 to 11 Years. Asia Pac J Public Health. 2016;28(5 Suppl):74S-84S. doi: 10.1177/1010539516641505.
17. Chiong R, Fan Z, Hu Z, Chiong F. Using an improved relative error support vector machine for body fat prediction. Comput Methods Programs Biomed. 2021; 198:105749. doi: 10.1016/j.cmpb.2020.105749.