Fuzzy Model for Prediction Childhood Obesity Cases Using the Generalized GFID3

Christian Suca, André Córdova, Abel Condori, Jordy Cayra y José Sulla

Nacional university of San Agustín, Professional School of Systems Engineering, Arequipa, Perú

csucavelando@gmail.com, andrepacord@gmail.com, acondoricas@unsa.edu.pe, jordd.perezz@gmail.com, josullato@gmail.com

Abstract. Obesity is cause of major chronic diseases, and therefore is one of the biggest health problems in the world, being the IMC the primary indicator used for prediction, however, the percentage of body fat is one of the factors associated with obesity problems, being these percentages different for men and women. The aim of this paper is to develop a new classified novel fuzzy model for men and women for the prediction obesity cases in people of aged 6-17 years. This paper studied the factors and indicators used as measures of prediction cases of obesity and developed a fuzzy model which can be used as a predictor of obesity in the medical field using percentiles weight, height and percentage of body fat (BF). The steps involved in this study are: a review of the factors of childhood obesity, pre-processing, data collection, classification and fuzzification of inputs and generating optimal rules based on the GFID3 tree. The results show that the developed fuzzy model gets accuracy of 83.65% and 76.13% for men and women for predicting obesity cases.

Keywords: Obesity, GFID3 Tree, Fuzzy Logic, IMC.

1. Introduction

The clinical condition known as obesity, currently assumes the characteristic of universal epidemic. That disease is a major cause of death and related diseases (also called co-morbidities). On the other hand, it is directly related to the magnitude of the clinical condition. WHO [6] indicates that the number of infants and children who are overweight increased from 32 million to 42 million in 2013, the WHO states that according to the tendency possibly this number will grow to 70 million in 2025. The definition of obesity, measurement, classification and treatment is not easily determined, find a system to evaluate and rank the level of obesity, and especially if it concerns the risks or prescription treatments is of great clinical interest [1]. The excess weight may come from muscle, bone, fat or water in the body, the presence of excess fat in the trunk or abdomen out of proportion with respect to total body fat is an independent predictor of risk factors and morbidity [19]. Nowadays multiple investigations are realized to classify a person on condition of overweight, obesity, etc., by using various anthropometric parameters.

When the interest is to determine the best therapeutic indication and treatment as well as the best appropriate mechanism for data analysis, it is necessary the search for a more accurate method to assess obesity and consequently the best treatment [1]. For this reason, the question arises, is it possible the implementation and use of a fuzzy model based on decision trees that can be used as a predictor of obesity cases by using a set of medical records growth of 6-17 people?

In this research, a fuzzy model for predicting of obesity cases is developed. This novel approach combines techniques of ID3 fuzzy decision tree and fuzzy logic for generating a model based on a set of rules as well as a model for predicting cases of obesity using as reference: age and weight percentiles and BF index which can be used as aid in the medical field. This model is classified for both men and women, using as reference to the fact that measures, ranges, as well as the percentage of fat have different values for men and women [1]. Therefore, the classified model is able to predict whether a person is prone to obesity cases more precisely and offer an alternative aid in the medical field. To relate the different variables of our database, it is used associative classification that is an integrated field of Association and Classification Rule Mining (ARM). The traditional ARM was designed using as reference that the elements are equally important and their presence or absence in the database is simply mentioned [18].

The paper is structured as follows. In section 2, related works, where investigations that were carried out previously is shown, section 3, the methods and tools used in the fuzzy model is shown, then in section 4, experiments and results that were obtained as well as the comparison with other articles that analyzed the same problem. Finally, in section 5 conclusions and future work, it is established the result that was obtained by the proposed model.

2. Related Works

In 2008 a new index measure was proposed as an alternative in decision-making for bariatric surgery with respect to their degree of obesity using as reference variables, BMI and body fat percentage in adults and using as reference the guidelines of WHO and expertise of the authors for generating rules [1]. In the global context of health, the body mass index (BMI) is considered as the main criterion for determining if a person has problems of obesity. Nevertheless, excess fat with respect to the percentage of body fat (% BF), it is actually the main detrimental factor in obesity disease that usually is overlooked.

In 2013 a fuzzy model for the evaluation of abdominal obesity (AO) and Cardio-metabolic risk (CMR) was presented by [2], on the basis of the best-known anthropometric measurement indicators, such as the body mass index (BMI), waist circumference (CC) and the relationship waist-height (RCEst).

In 2014 the fuzzy medical diagnosis system was made [3] to support and help to evaluate patients with chest pain according to the degree of obesity. In that research, the body mass index (BMI) was fuzzified. By the same year the author in [4] proposed a new fuzzy decision tree which uses information entropy (abbreviated as GFID3), which improved the shortcomings of existing algorithms based on fuzzy decision trees, because they do not take into account the impact of nonlinearity, so they are unable to integrate the selection of extended attributes. In 2015 [5] realized a research in children 4 to 6, that research takes into account the body mass index and level of physical activity for analysis of obesity in children.

3. Methods and Tools

The dataset used is a set of 5962 medical records of people between 6-17 years, provided by educational institutions in Brazil. The selected parameters that were used for training are shown in Table 1.

Table 1. Input parameters.

Parameters	Description
Height percentile	Classification index of height by age based on percentiles (0-100).
Weight Percentile.	Classification index of weight by age based on percentiles (0-100).
BF	Body Fat

After having done that, BMI was computed according to the clinical guideline established by [6]; this computing was performed with respect to height and weight percentiles and gender. After calculated BMI with respect to children and adolescents, the percentile values of BMI and age are shown in Figure 1. Those values were provided by the Centers for Control and Prevention of Diseases (CDC). BMI by age for children in percentile values are shown in Fig. 1 and 2.

Fig. 1. Body mass index in boys based on percentiles [6]

Percentiles are the indicator used most often to assess the size. The percentile indicates the relative position of the child's BMI among children of the same sex and age. The weight level categories of BMI by age and its corresponding percentiles are shown in Table 2.

Table 2. Clinical Guidelines for identifying obesity in children [6].

Classification	Percentile IMC
Underweight	lower than 5
Normal	greater than 5 and lower than 85
Overweight	greater than 85 and lower than 95
Obesity	greater than 95

BF is considered a better index to assess obesity in comparison with BMI because this one only works as an indication of excess of weight and BF is an indicative of excess of fat mass [1] [7]. Therefore, BF has been fully exploited in this research. Values for girls and boys are shown in Tables 2 and 3 respectively.

Table 3. Clinical guidelines for assessing the percentage of fat in boys [7].

Classification	Percentages of fat
Very Low	greater than 0 y lower than 5.5
Low	greater than 5.5 and lower than 11.5
Optimal range	greater than 11.5 and lower than 21
Moderately High	greater than 21 and lower than 26
High	greater than 26 and lower than 31
Very High	greater than 31 and lower than 100

Table 4. Clinical guidelines for assessing the percentage of fat in girls [7].

Classification	Percentages of fat
Very Low	greater than 0 and lower than 11
Low	greater than 11 and lower than 15
Optimal range	greater than 15 and lower than 23.5
Moderately High	greater than 23.5 and lower than 31.5
High	greater than 31.5 and lower than 34.5
Very High	greater than 31 and lower than 100

To introduce a coherent mechanism and tackle obesity, in this paper was considered both weight and height percentile and BF to present a new fuzzy index for obesity prediction in children by using the fuzzy logic.

4. Experiments and Results

It was considered 4 steps in this paper, they are: data pre-processing, fuzzification, generation of optimal rules, application of Mamdani system and classification of outcomes.

4.1 Pre-Processing Data

In the initial stage, was available a database that had a total of 14 attributes including age, sex, body mass index, physical exertion variables, skinfold thickness, etc. According to [1] Crisp values were established to delimit the limits and ranges of the attributes BMI and BF. Nevertheless, these measures are appropriate for adults. Therefore, it is necessary to establish according to experts what are the measurement attributes which are critical factors for obesity prediction [7] [1] The author in [1] proposed a new classification index of obesity for adults using as reference BMI and percentage of fat, with reference to the results given by [1] was taken as inputs to the system, the percentage of body fat as well as weight and height percentiles, since the proposed model is aimed at people aged 6-17. The values of fat percentage (BF) have Crisp different values for men and women, so it was necessary a classification of data records in men and women. Then it was necessary the conversion of the attributes of BMI, weight and height to a percentile, because as mention WHO [8] a basic measure for controlling obesity is the percentile of BMI because this measure allows us to establish a percentage and compare it with other assessments made other children and check if it is in an acceptable range. For doing that, a base template was used to convert BMI into percentiles given by [9] The next step was a classification of data in men and women as well as the calculation of trapezoidal values for establishing membership function by using crisp values and the average of each of the records with respect to each crisp limit as well as the standard deviation to be used in the fuzzification stage.

4.2 Fuzzification

After preprocessing data, it was needed to have prepared ranges and classes for each one of these attributes, each one were based on expert guidance such as WHO and other research papers on Childhood Obesity. Crisp values and fuzzy limits based on weight and height percentiles were obtained of WHO [10] that provides a series of graphs and tables of percentiles for both boys and girls, being classes: Low weight, normal weight, overweight, and obesity. As for the percentage of fat according to [7] which establishes stated according to skinfold the result of fat mass, lean mass and body fat percentage for children being this categories: Very low, low, normal, moderate, and Very High. Then it was also considered the sorting outlet, this is important because initial data or pre-training that will serve in the step of generating rules is needed ranges and classes were used as input for the Mamdani fuzzy logic system. The membership function which was used is the trapezoidal function because according to [1] is the most suited, trapezoidal values and ranges were achieved based on the average and standard deviations of each range to get the four parameters of each trapezoid. Trapezoidal membership function is shown in equation 1, where x is the input attribute to be fuzzified, a, b, c, d are the ranges of the trapezoid.

$$f(x,a,b,c) = \left\{ \begin{array}{cc} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & d \le x \end{array} \right\} (1)$$

Finally, the results of each attribute with its fuzzy values or graduated one were achieved where each attribute can belong more than one category. In the following Figures 3, 4 are shown membership functions of body fat attribute.

,

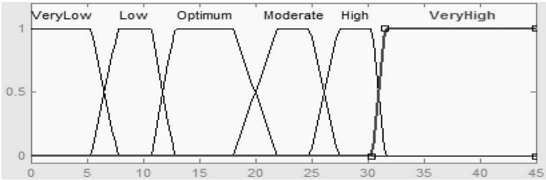

Fig. 2. BF index in boys.

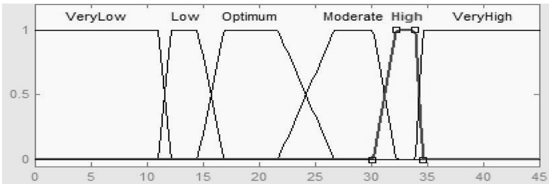

Fig. 3. BF index in girls.

4.3 Generation of Optimal rules

For this stage, some algorithms have advantages in terms of robustness, but also have disadvantages as falling into local optima. Other techniques require large computational effort to process it, and some lack simplification methods to obtain other rules with high precision, there are neural networks [11], cognitive maps [12], decision trees [13, 14, 15] or genetic algorithms [16], all generalized one to work with fuzzy data.

GFID3 Fuzzy Tree

In the recursive construction of tree, it is adopted strategy 'divide and rule'. ID3 calculates the fuzzy partition entropy based on the attribute and then selects the attribute with minimal partition fuzzy entropy, the fuzzy partition entropy is an extension of the entropy of partition Crisp algorithm ID3 [17].

Let $\Omega = \{1,2,\dots,n\}$ a training data, $A = \{A_1 A_2,\dots,A_m\}$ a set of attributes where $\{A_{i1}, A_{i2},\dots,A_{iki}\}$ is the range of each attribute, in the GFID3 tree with fuzzy values, decision nodes are considered as fuzzy sets in the fuzzy partition, entropy in a non-leaf node D is given by $FE(D, A_i)$.

$$FE(D, A_i) = \sum_{j=1}^{ki} \frac{m_{ij}}{mi} E\left(D, A_{ij}\right) \qquad (2)$$

$$E\big(D, A_{ij}\big) = -\sum_{k=1}^{ni} \frac{m_{ijk}}{m_{\overline{ij}}} log_2 \frac{m_{ijk}}{m_{\overline{ij}}} \qquad (3)$$

Where $m_{ij} = M(A \cap A_{ij})$,$m_i = \sum_{j=1}^{k} m_{ij}$, $m_{ijk} = M(D \cap A_{ij} \cap C_k)$,$\overline{m_{ij}} = \sum_{k=1}^{n} m_{ijk}$

One of the characteristics fuzzy of decision tree ID3 based on generalization of entropy (GFID3) proposed by [17] is that it can be applied to both data set with fuzzy attribute values and numeric and continuous attributes.

The selected method was the generalized algorithm based on fuzzy decision tree GFID3 given by [17]. In addition to considering the good results provided by this technique, it is necessary to consider especially the speed with which can provide the rules. The rules obtained, and the accuracy rates for each output class are shown in Tables 4 and 5.

Table 5. Fuzzy rules for Girls.

Number	Rules	Accuracy
1	IF Weight is Normal AND Height is Very High AND BF is Very High THEN Low Weight	1.00
2	IF Weight is Overweight AND BF is Optimal AND Height is Low THEN Obesity	1.00
3	IF Weight is Overweight AND BF is High AND Height IS Low THEN Overweight	1.00
4	IF Weight is Overweight AND BF is Very High THEN Obesity	1.00
5	IF Weight is Normal AND Height is Low THEN Overweight	0.98
6	IF Weight is Normal AND Height is Normal AND BF IS Very Low THEN Low Weight	0.90
7	IF Weight is Overweight AND BF is Low AND Height IS Alta THEN Normal	0.90
8	IF Weight is Normal AND Height is Very High AND BFIS Moderate THEN Low Weight	0.88
9	IF Weight is Overweight AND BF is Low AND Height IS Normal THEN Normal	0.86
10	IF Weight is Overweight AND BF is Moderate THEN Overweight	0.84

Table 6. Fuzzy rules for boys.

Number	Rules	Accuracy

1	IF Weight is Low Weight AND Height is Normal AND BF is moderate THEN Low Weight	1.00
2	IF Weight is Normal AND Height is Low AND BF is Optimal THEN Normal	1.00
3	IF Weight is Normal AND Height is Low AND BF is High THEN Overweight	1.00
4	IF Weight is Normal AND Height IS Normal AND BF is moderate THEN Overweight	1.00
5	IF Weight is Normal AND Height is Normal AND BF IS High THEN Overweight	1.00
6	IF Weight is Overweight AND BF is Very Low AND Height is Normal THEN Normal	1.00
7	IF Weight is Overweight AND BF is Very Low AND Height is Over Height THEN Normal	1.00
8	IF Weight is Overweight AND BF is Very Low AND Height is High THEN Low Weight	1.00
9	IF Weight IS Overweight AND BF is Very High AND Height is Low THEN Obesity	1.00
10	IF Weight IS Overweight AND BF is Very High AND Height is High THEN Overweight	1.00

Finally, fuzzy rule sets were inserted into fuzzy system, being such rules which will become part of the result or output of the fuzzy system.

4.4 Classification and Results

After providing training dataset with 5962 records in total, 20% training dataset were used to train GFID3 tree whereas total dataset was used to test stage.

As for the results of comparison, the aim was to assess the accuracy category of each rule as well as the outcomes. In Tables 6 y 7 is shown some outcomes of 10 first records for men and women respectively.

Table 6. Comparison of outcomes for boys.

Height Percentile	Weight Percentile	BF	Outcome	Classification of model	Expert outcome	Expert classifications
56.56	37.03	11.9	38.3	Normal	27.273	Normal
83.45	54.57	10.1	43.40	Normal	24.03	Normal
52.75	55.81	11.5	43.85	Normal	60.84	Normal
86.76	96.28	27.8	97.54	Obesity	96.21	Obesity
99.93	99.97	33.5	97.54	Obesity	99.71	Obesity
97.37	99.41	28.6	97.54	Obesity	98.70	Obesity
87.30	99.28	30.2	97.541	Obesity	99.13	Obesity
98.58	97.00	18.5	85.19	Overweight	91.24	Overweight
96.40	96.66	17.8	85.31	Overweight	93.15	Overweight
97.98	97.14	16.7	85.43	Overweight	92.35	Overweight

Table 7. Comparison of outcomes for girls.

Height Percentile	Weight Percentile	BF	Outcome	Classification of model	Expert outcome	Expert classification
82.40	30.15	13.7	7.38	Normal	1.69	Low Weight
78.08	58.84	11.9	45.29	Normal	37.25	Normal
84.90	54.29	15.4	43.82	Normal	19.87	Normal
100.00.	100.00	36.2	97.53	Obesity	99.82	Obesity
86.53	99.28	30.7	97.53	Obesity	99.38	Obesity
98.73	99.60	33.8	97.53	Obesity	98.92	Obesity
98.39	99.57	24.6	97.53	Obesity	98.92	Obesity
70.34	90.57	30.7	86.08	Overweight	94.02	Overweight

The classification outcomes using 2935 boy records and 3021 girl records is shown in percentages in the Tables 8, 9.

Table 8. Classification of outcomes using boy records.

Boys	Accuracy
Classified instances correctly	83.65
Classified instances incorrectly	16.35

Table 10. Classification of outcomes using girl records.

Girls	Accuracy
Classified instances correctly	76.13
Classified instances incorrectly	23.87

4.5 Discussions

According to the fuzzy system and the classification results with the fuzzy algorithm based on decision trees GFID3, it is clear that the fuzzy system is better in terms of classification giving an average accuracy of rules 82.79% for children and a percentage of 76.37% accuracy for girls. In comparison the research which was made by Miyahira and fuzzy index based on BMI and BF proposed for the prediction of cases of obesity in adults, this research took as input into fuzzy system, percentile values of BMI weight, height and % BF, considering the distribution of these in the age range of 2-20 years and generating the most optimal rules by using 5643 medical records and application of GFID3 algorithm that has proved to be more robust than fuzzy decision tree based on ID3 and therefore right in the medical field..

5. Conclusion and future work

The paper presents a fuzzy index for predicting cases of obesity in both men and women taking as inputs the percentiles of weight, height and% BF, based on logic and fuzzy sets. The result obtained are a set of rules and a fuzzy model for determining whether a person is prone to obesity problems for your medical records of weight, height and% BF obesity, these rules were generated by fuzzy inputs and the GFID3 tree that has shown great accuracy against the proposed variants of ID3. The developed fuzzy index can be used in the medical field for the prediction of cases of obesity, however, it is not intended to replace or substitute assessment metrics which are used by doctors with experience. On the other hand, it should be seen as a support tool for the prediction of cases of obesity in people aged 6-17 years. As future work, this research will be extended for adults, optimizing the set of rules generated by an evolutionary algorithm such as genetic algorithms to compare with a neural network and test and validate their findings in the medical field.

References

1. S. A. Miyahira, J. L. M. C. de Azevedo, and E. Araújo, "Fuzzy obesity index (MAFO I) for obesity evaluation and bariatric surgery indication.," J. Transl. Med., vol. 9, no. 1, p. 134, 2011.
2. Nawarycz, T.; Pytel, K.; Drygas, W.; Gazicki-Lipman, M.; Ostrowska-Nawarycz, L., "A fuzzy logic approach to the evaluation of health risks associated with obesity," in Computer Science and Information Systems (FedCSIS), 2013.
3. Orsi, T.; Araujo, E.; Simoes, R., "Fuzzy chest pain assessment for unstable angina based on Braunwald symptomatic and obesity clinical conditions," in Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on, vol., no., pp.1076-1082, 6-11 July 2014.
4. C. Jin, F. Li, and Y. Li, "A generalized fuzzy ID3 algorithm using generalized information entropy," Knowledge-Based Syst., vol. 64, pp. 13– 21, Jul. 2014.
5. M. Khanna, D. N. K. Srinath, and D. J. K. Mendiratta, "The Study of Obesity in Children Using Fuzzy Logic," IJITR, vol. 3, no. 1, 2015.
6. WHO, (2007-2009). Comisión para acabar con la Obesity infantil: Datos y cifras Over Obesity infantil Publishing. Recuperado de http://www.who.int/end-childhood-obesity/facts/es/.
7. P. – C. P. Y. V. – A. J. Hall –López J, Ochoa –Martínez P, Monreal – Ortiz L, "Porcentaje de grasa corporal en niños de edad escolar," vol. II, no. 1, pp. 13–17, 1988.
8. WHO, Centers for Disease Control and Prevention CDC, 2015.
9. Luis F. Romero, Tabla de Crecimiento (Control de Percentiles), 2009 Disponible en: http://www.ac.uma.es/~felipe/personal.html
10. OMS, CDC growth charts to monitor growth for children age 2 years and older in the U.S. available in http://www.cdc.gov/growthcharts/clinical_charts.html.
11. Jang, J.-S.R., "ANFIS: adaptive-network-based fuzzy inference system," in Systems, Man and Cybernetics, IEEE Transactions on, vol.23, no.3, pp.665-685, May/Jun 1993
12. B. Kosko, "Fuzzy cognitive maps," Int. J. Man. Mach. Stud., vol. 24, no. 1, pp. 65–75, Jan. 1986.
13. Koen-Myung Lee; Kyung-Mi Lee; Jee-Hyong Lee; Hyung Lee-Kwang, "A fuzzy decision tree induction method for fuzzy data," in Fuzzy Systems Conference Proceedings, 1999.
14. Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," Fuzzy Sets Syst., vol. 69, no.2, pp. 125–139, Jan. 1995.
15. Janikow, C.Z., "Fuzzy decision trees: issues and methods," in Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol.28, no.1, pp.1-14, Feb 1998
16. Y. Yuan and H. Zhuang, "A genetic algorithm for generating fuzzy classification rules," Fuzzy Sets Syst., vol. 84, no. 1, pp. 1–19, Nov. 1996.
17. [17] C. Jin, F. Li, and Y. Li, "A generalized fuzzy ID3 algorithm using generalized information entropy," Knowledge-Based Syst., vol. 64, pp. 13– 21, Jul. 2014.
18. [18] S. Soni and O. P. Vyas, "Fuzzy Weighted Associative Classifier: A Predictive Technique for Health Care DataMining," Int. J. Comput. Sci. Eng. Inf. Technol., vol. 2, no. 1, pp. 11–22, 2012.
19. [19] S. Bouharati, M. Bounechada, A. Djoudi, D. Harzallah, F. Alleg, and H. Benamrani, "Prevention of Obesity using Artificial Intelligence Techniques," vol. 1, no. 9, pp. 146–150, 2012.