International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Aggrandizing the Accuracy of Body Fat Percentage by Stratification using Decision Tree

J. Grace Hannah, D. Gladis

Abstract*: Bariatrics is the branch of science which deals with obesity and its related surgical procedures. A person's physical indolence, unhealthy food habits and genetic constitution emanates as the fons et origo of health gremlins. Multifaceted indagations have been worked on the diverse and heterogenous obstinate concerns caused due to obesity. Anatomization of body fat percentage has become a rudimentary regimen for every individual to be done in a fastidious manner. The whilom work anent body fat percentage entailed Body Mass Index (BMI) with respect to age and gender of a person. The anatomical conformation of an individual unraveling the fat constitution and the muscle tissue composition is not computationally enumerated using BMI. Thus, the formula using BMI dempers the veracity for a person having more muscle mass than fat mass and speciously vitiates the fat percentage of that person. The proposed novel formula is analyzed by cross-validated classification model using decision tree, and is effectuated by implementing information gain. This accentuates the coherence, efficacy and accuracy of the derived body fat percentage for a person. The Ethical Committee approval for this study has been obtained from the Institutional Ethics Committee, Madras Medical College, Chennai. The empirical study has been simulated using Matlab and the results have been successfully acquired in the GUI mode.

***Keywords* : Obesity, Body Fat Percentage, BMI, Decision Tree, Matlab GUI**

I. INTRODUCTION

An augmented nodus under the radar which has become a cardinal thread of discussion, and which has raised a lot of awareness in various parts of the world is "Obesity". In recent times, surveys have shown that nearly thirty percent of the entire population is obese, and almost three million reportedly suffer from concomitant obesity lurgies including heart diseases, cancer and diabetes. Most of the world's population lives in countries where overweight and obesity kills more people than underweight. The study organized by WHO also showed that 41 million children under the age of 5 were overweight or obese [5], [19],[20]. A neoteric and unabating realm in the health industry is the advancement in bariatric sector [20]. The age at which a person becomes obese metamorphoses the ability of the individual to lose calories. The energy imbalance between the ingested calories and the calories expended is considered as the outset of obesity and uncurbed fat [19],[20]. Obese individuals can have great difficulty losing weight. The uncurbed energy and White Adipose Tissues (WAT) can cause serious detriments like Cushing's syndrome, Hypothyroidism, Neurologic disturbances and various other health impediments [9], [19], [20]. Though dieting and exercise are solutions for weight reduction, studies show that only the size of the fat cells are pruned, instead of annihilating them completely [4], [19], [20]. To anatomize the growth of adipose cells in the human body a meticulous analysis about the White Adipose Tissue (WAT) and Brown Adipose Tissue (BAT) is done [20].

Brown Adipose Tissue (BAT) are an important integrant which proselytizes the excessive fat in the body by incinerating them [6], [7], [19], [20]. They comprise of several lipid droplets, with ginormous quantum of iron. On the other hand, WAT consists of a single lipid droplet, less mitochondria, and are regarded as the unbridled, unseared body fat which usually leads to obesity and other diseases [8], [19], [20]. In children, the excess calories and WAT are converted into new fat cells (hyperplasic obesity) [1], [2], [19], [20], while the nimiety of calories consumed by an adult proliferates the existing fat cells (hypertrophic obesity) [4], [19], [20]. The body fat percentage is a primary factor that connotes if a person is healthy or not. Body fat is divided into two categories such as the subcutaneous fat and the visceral fat [11], [13], [19], [20]. Subcutaneous fat is that which is located beneath the skin [11], [19], [20], while the visceral fat is located inside the peritoneal cavity surrounding the peritoneal organs [12], [19], [20]. The various approaches for treating obesity incorporate studies of hormones like Leptin and Ghrelin [20]. These hormones are secretions of the fat cells in the body. Leptin is concomitant to the reproductive function, and Ghrelin stimulates the pituitary gland to release growth hormones. Hyperlipidemia and proliferation of fat are often analogized as they correlate and aid in procuring better cognizance of obesity disorders [20]. The previous work of study involved the assessment of BMI, but for an athlete or a healthy person with more muscle mass, the above formula using BMI can lead to fallacious results. This paper aims to effectuate and proliferate the efficiency and precision of the derived body fat formula using the Lipid Profile and Bio-Electric Impedance method from an individual. The classification of the different classes such as the obese and non-obese category in male and female using decision tree analysis and the respective information gain for each class are computed. The paper is organized as follows: section II gives a brief study of the techniques implemented; section III annotates the reinforcement of the aggrandized accuracy using the novel formula for body fat percentage. Section IV presents a schematic overview of the results obtained from the decision tree analysis, and section V concludes the work done with the future work that can be implemented.

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
J. Grace Hannah*Research Scholar, Department of Computer Science, Presidency College, Affiliated to The University of Madras, Chennai, India, Email: hannahgrace161289@gmail.com
Dr. D. Gladis, Principal, Bharathi Women's College.Chennai

II. TECHNIQUES USED

The study population includes adult patients who were obese, hyperlipidemic and faced other problems due to unbridled disintegration of fat cells. The data samples were obtained from Rajiv Gandhi Government General hospital, Park Town, Chennai. The body fat percentage is computed by formulating the values obtained from the blood tests and the bioelectric impedance method for a person [19], [20]. Pregnant women, patients having bleeding disorders, anaemic and suffer from any other serious ailments were excluded from the study [19], [20].

The lipid profile test involved the sample blood drawn from patients who fasted for 9-12 hours. The lipid levels are checked for factors such as Total Cholesterol (TC), High-density Lipoprotein (HDL), Low density Lipoprotein (LDL) and Triglycerides (TG), which are then centrifuged and then combined with the Cfas lipid reagent along with the necessary preci-controls to acquire the values of each of the factors. [20]

The Bio - Electric Impedance Analysis (BIA) is a non- invasive method. It employs a technology which is extensively used, and aids in providing impeccable analysis of the body composition for an individual [15], [19], [20]. It operates at a temperature humidity of 5 to 35 degree Celsius with no corrosive gas in the surrounding air. Bioelectric Impedance expounds as the opposition of a conductor to the flow of an alternating current, and consists of two constituents: resistance (R) and reactance (Xc) [17], [18], [19], [20]. The Resistance has a low frequency of 40-50 kHz and is the radical opposition of the conductor [18], [20]. Reactance is the additional opposition or the ephemeral storage of the electric charge by a condenser. The lipid components of the membranes of the Body Cell Mass (BCM) behave as capacitors and reduce the flow of intracellular ions [18], [19]. Despite a general perception that BIA estimates body fat, the technology actually determines the electrical impedance of body tissues and adiposity [16], [19]. The BIA quantifies the resistance of body tissues through the flow of small electrical signals through the ions present in the fluid content of a human body. The device uses a pair of electrodes through which a low level imperceptible electric current flows and divulges the estimates of water content and body fat in the blood, tissues and bones. The fat in the body is analyzed depending on where they are distributed. The electric current which passes through the body is at differential rate depending on the body composition [18], [19], [20]. Hence there is a direct relationship between the concentrations of the ions and electric conductivity, and an indirect relationship exists between the ion concentration and the resistance [18], [19]. The BIA enumerates and itemizes the body facets such as visceral fat percentage, subcutaneous fat percentage and segmental subcutaneous fat percentage, age of the individual cognate to their fat percentage, Body Mass Index (BMI), muscle mass percentage, its corresponding segmental muscle mass percentage and the total body fat percentage [19], [20]. The BIA is currently used in diverse settings including private clinics, hospitals and across a spectrum of ages, body weights and disease states [19], [20].

A. Modus Operandi

The dataset procured from the blood samples and the BIA method is simulated in MATLAB to peruse the precision, coherence of the derived body fat formula, and to catalogue the datasets germane to the body fat percentage and cholesterol levels of an individual. The datasets go through a nexus of calibrations before the denouement stratification into a decision tree model. The inceptive phase entails an aleatoric permutation to bifurcate the dataset into testing and training dossiers. Pre-processing of data is a rudimentary aspect to curtail noise and coarct the data without any attenuation. Thus, the segregated datasets are discretized and binned as a part of pre-processing based on the prevenient cardinal number of bins selected. The learning approach recursively divides the training data into buckets of homogenous members through the most discriminative dividing criteria [24]. The ensuing phase involves the computation of information gain for all the attributes, which leads to the sequent selection of the optimum features from the dataset. The two most pivotal measures for attribute selection are information gain and gain ratio. In this paper, the attribute selection of the best features is steered by information gain, and the splitting criterion is implemented by gini index. This yields the divergence between the probability distributions of the target attribute's value. The conclusive taxonomy is contrived by using the fit classification tree with binary splits in MATLAB. The Decision Tree classifiers are comprehensively exerted for accurate medical prognosis. A decision tree is a chronological representation of nodes. It is deployed using a supervised learning algorithm that follows greedy approach to classify the unknown data samples. The tree pruning is accomplished by setting the 'prune' function.

If Prune is 'on', then fitctree grows the classification tree and estimates the optimal sequence of pruned subtrees, but does not prune the classification tree. Otherwise, fitctree grows the classification tree without estimating the optimal sequence of pruned subtrees.

III. PERFORMANCE ANALYSIS

The existing body fat formula is given as:

[(1.2*BMI) + (0.23*Age) – (10.8*Gender) – 5.4] [14], [19], [20]

Though the above formula has been widely accepted as the standard measure to analyze body fat percentage, it has been observed that for a person who is muscular and has more muscle mass, the BMI increases considerably, thereby placing him/her in the morbidly obese or overweight category [19], [20]. Therefore, to avoid fallacious calculations, the below formula involving the total Cholesterol (TC) from the lipid profile, subcutaneous and visceral fat from the BIA, age pertaining to the fat accumulation in the body and gender of the person is taken into consideration. The optimal feasible formula thus obtained after performing regression is as follows [19], [20]:

Body fat % = [(0.5613 *(SF+VF)) +(0.0572 (Age(M) -Age) + (0.0290 * TC)- (10.889 *Gen) + 10.0822] *(1)*

Where SF & VF signifies the subcutaneous and the visceral fat, age (M) is the age of the person with respect to the body fat percentage, gender (Gen) takes the value of 0 for female and 1 for male [19], [20].

A performance analysis has been performed to vindicate and augment the derived formula which has been subsumed in the decision tree model to procure the precision rate for each of the classified classes. The classification analysis has rendered an accuracy of 96% and 84% for all classes in the training phase and testing phase respectively.

IV. RESULT OBTAINED

The dataset incorporated into MATLAB provides better cognizance of classification, and has yielded the below results in GUI.

Figure 1 elucidates the discretized dataset with the display of information gain for each attribute.

Figure 2 shows the best feature selection using information gain.

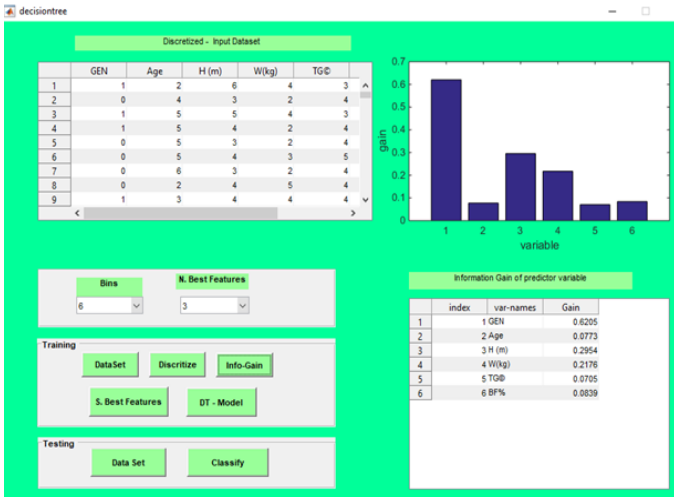

Fig 1 : Discretized Dataset with Information gain of attributes

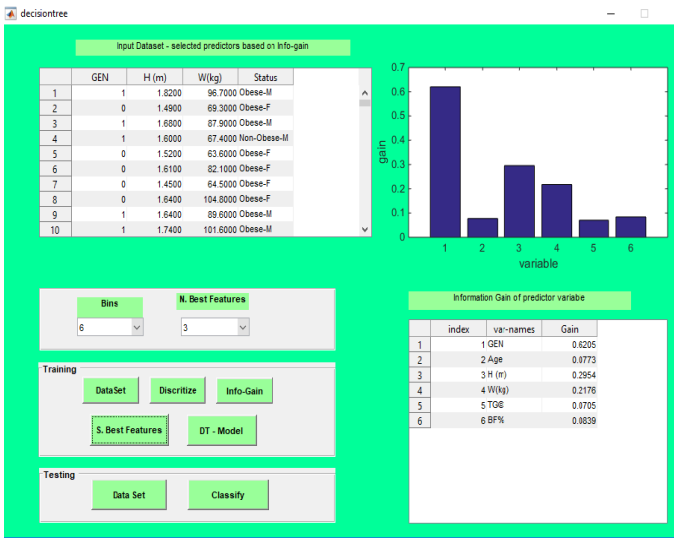

Fig 2: Attribute Selection using Information Gain

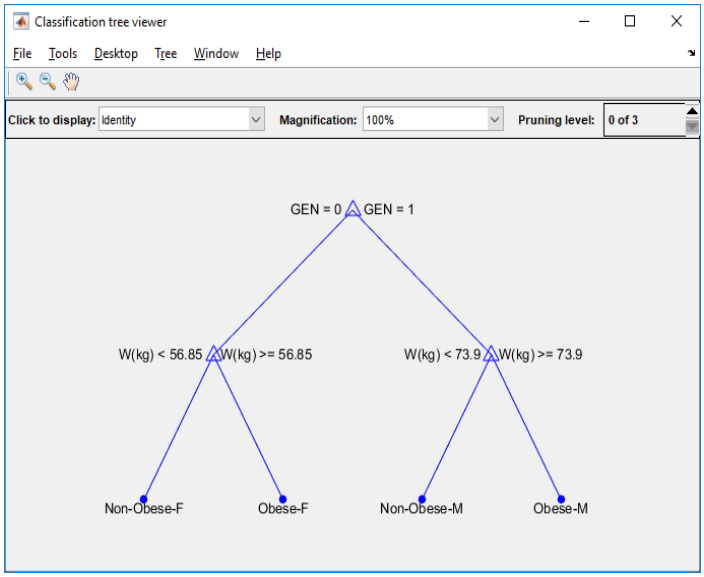

Fig 3 : Cross validated decision tree model in Matlab

V. CONCLUSION

This study has bolstered the computation of the novel body fat percentage by explicating the classification methodology in the form of a decision tree. The computation of the subcutaneous and visceral values has contributed to fathomless intellect of an individual's intramural and over the skin fat depositions. The novel formula aids to proselytize and overcome the gremlins encountered by using BMI, and has yielded a mean error rate of less than 0.4%. Thus, making it an unambiguous and unswerving measure to be used in the future to meliorate the health milestones. The simulation in MATLAB has asseverated the classification precision of each class by entailing information gain, which has vouchsafed the prominence of the attributes in the dataset. This has capacitated the cogent panacea of the gordian knot by generating very less error rates in the training and testing phase of the decision tree classification.

REFERENCES

1. A D Sniderman, J Mailloux, M Fortier, C Huot, C Rhéaume and A Tchernof, "Hypertrophy and hyperplasia of abdominal adipose tissues in women", International Journal of Obesity (2008) 32, 283–291; doi:10.1038/sj.ijo.0803708; published online 28 August 2007.
2. Carolina T. Mendes- dos-Santos, Sofia H. V. de Lemos-Marini, Maria Tereza M. Baptista, Gil Guerra-Junior, Maricilda P. De-Mello, Maria Fernanda V. M. Paulino, André M. Morcillo, "Normalization of height and excess body fat in children with salt-wasting 21-hydroxylase deficiency", J. Pediatr, (Rio J.) vol.87 no.3 Porto Alegre May/June 2011
3. https://www.niddk.nih.gov/health-information/weight-management/bariatric-surgery/ definition-facts, Definition & Facts for Bariatric Surgery.
4. Lester b. Salans, Samuel W. Cushman, and Rodger E. Weismann, "Studies of Human Adipose Tissue, Adipose cell size and Number in non obese and obese patients", The Journal of Clinical Investigation, Volume 52.
5. http://www.who.int/mediacentre/factsheets/fs311/en/obesity and overweight
6. David A. Bernlohr, Anne E. Jenkins and Assumpta A. Bennaars, "Adipose tissue and lipid metabolism", Elsevier Science, 2002.

7. Christoph H. Saely, Kathrin Geiger, Heinz Drexel, “Brown versus White Adipose Tissue: A Mini-Review”, Gerontology 2012
8. ClÁudia M. Oller Do Nascimento, Eliane B. Ribeiro and Lila M. Oyama, “Metabolism and secretory function of white adipose tissue: Effect of dietary fat”, Anais da Academia Brasileira de Ciências (2009) 81(3): 453-466.
9. Health risks of being Overweight, U.S. Department of Health and Human Services, October 2007.
10. Rômulo A. Fernandes, Clara S. C. Rosa, Camila Buonani, Arli R. de Oliveira, Ismael F. Freitas Júnior, “The use of bioelectrical impedance to detect excess visceral and subcutaneous fat”, Journal de Pediatria, August 2007.
11. Fredrik Toss, “Body fat distribution, inflammation and cardiovascular disease”, ISSN: 0346-6612; Sweden, 2011
12. Chun-Hao Chen, Yu-Yawn Chen, Chih-Lin Chuang, Li-Ming Chiang, Shu-Min Chiao and Kuen-Chang Hsieh, “The study of Anthropometric estimates in the Visceral fat of healthy individuals”, Nutrition Journal, 2014.
13. Body Fat Percentage: A Complete Guide to Evaluation and Measurement, Leigh Peele, 2010
14. Paul Deurenberg , Kalpana Bhaskaran and Petrina Lim Kim Lian, “Singaporean Chinese adolescents have more subcutaneous adipose tissue than Dutch Caucasians of the same age and body mass index”, Asia Pacific J Clin Nutr 2003;12 (3):261-265
15. Sun SS, Chumlea WC, Heymsfield SB, “Development of Bioelectric impedance analysis prediction equations for body composition with the use of multicomponent model for use in epidemlogic surveys”, American Society for clinical Nutrition.
16. DK Key, Bosaeus “Body Composition estimated by Bioelectric impedance in the Swedish elderly - Development of population – based prediction equation and reference values of fat – free mass and body fat for 70 – 75 year olds” , European Journal of Clinical Nutrition.
17. Linda L.D. Zhong, Wai Kun et al, “The combination effects of body acupuncture and auricular acupressure compared to sham acupuncture for body weight control: Study protocol for a randomized controlled trial”, BioMed Central.
18. Mahshjid Dehgan et al, “Is Bioelectric Impedance accurate for use in large epidemilogical studies”, BioMed Central, Nutritional Journal, 2008
19. J.Grace Hannah, D. Gladis, “A Statistical Indagation of Body Fat Percentage: A Sift Cognitive Correlation of Lipid Data and Bio-Electric Impedance Analysis in Humans”, International Journal of Advances in Science Engineering & Technology, ISSN: 2321-8991, Vol. 6, Issue-2, April 2018
20. J.Grace Hannah, D. Gladis, “A cognitive study of Body Fat Percentage in Humans by Anatomization of Lipid Profile, Bio-Electric Impedance and Clustering using K-Means Algorithm”, International Journal of Engineering & Technology, 7(2.33)(2018) 835-838.
21. Classification using Decision Trees, Winter School on “Data Mining Techniques and Tools for Knowledge Discovery in Agricultural Datasets”,
22. “Classification Using Decision Tree Approach towards Information Retrieval Keywords Techniques and a Data Mining Implementation Using WEKA Data Set”, K.F.Bindhia et al, International Journal of Pure and Applied Mathematics, Vol 116, No.22 2017, 19-29, ISSN: 1311-8080
23. “Analysis of Various Decision Tree Algorithms for Classification in Data Mining”, Bhumika Gupta et al, International Journal of Computer Applications (0975 -8887), Vol 163-No.8, April 2017
24. “Theoretical Study of Decision Tree Algorithms to Identify Pivotal Factors for Performance Improvement : A Review”, Amita Sharma et al, International Journal of Computer Applications (0975 -8887) Vol 141-No.14, May 2016.

AUTHORS PROFILE

J. Grace hannah is a research scholar pursuing her Ph.D in the Department of Computer Science, Presidency College, Affiliated with The University of Madras. She has more than eight research papers to her credit, and has participated in various National and International conferences.

Dr. D. Gladis is the Principal of Bharathi Women’s College, Chennai. She has headed the Department of Computer Science & Research Institute in Presidency College, Chennai. She has guided more than fifteen research scholars under the M.Phil and Ph.D courses and has been an active researcher. Her areas of interests are Neural Networks, Image Processing, Data Mining to name a few.