Article - Human and Animal Health

Application of Receiver Operating Characteristics (ROC) on the Prediction of Obesity

Mohammad Khubeb Siddiqui[1]*
https://orcid.org/0000-0001-6699-6216

Ruben Morales-Menendez[1]
https://orcid.org/0000-0003-0498-1566

Sultan Ahmad[2]
https://orcid.org/0000-0002-3198-7974

[1]Tecnologico de Monterrey, School of Engineering and Sciences, Monterrey, NL, México; [2]Prince Sattam Bin Abdulaziz University, College of Computer Engineering and Sciences, Department of Computer Science, Alkharj, Saudi Arabia.

*Correspondence: khubeb@tec.mx Tel.: +52-5522466170

HIGHLIGHTS

- Data mining technique applied on Obesity data set.

- Predictive analysis is done on each Body Mass Index (BMI) category.

- Our results show that at each BMI category patient has to be careful.

- Prevention is must at BMI ≥ 25, as Area Under Curve (AUC) is 99%.

Abstract: Obesity is the most common chronic disease, due to its ignorance in society. It gives birth to other diseases such as endocrine. The objective of this research is to analyze the different trends of each BMI category and predict its related serious consequences. Data mining based Support Vector Machine (SVM) technique has been applied for this and the accuracy of each BMI category has been calculated using Receiver Operating Characteristics (ROC), which is an effective method and potentially applied to medical data sets. The Area Under Curve (AUC) of ROC and predictive accuracy have been calculated for each classified BMI category. Our analysis shows interesting results and it is found that BMI ≥ 25 has the highest AUC and Predictive accuracy compares to other BMI, which claims a good rank of performance. From our trends, it has been explored that at each BMI precaution is mandatory even if the BMI < 18.5 and at ideal BMI too. Development of effective awareness, early monitoring and interventions can prevent its harmful effects on health.

Keywords: data mining; support vector machine; receiver operating characteristics; area under curve; body mass index; obesity.

INTRODUCTION

Obesity is a chronic disease that simply means being overweight. According to the medical dictionary, it is an accumulation of excessive fat on the human body which results in impaired health [1,2]. Typically, obesity occurs due to the imbalance of calories in our body, as we do not burn these extra calories due to the ignorance, as a result obesity gradually occurs in our body [3]. The prevalence of obesity is growing

rapidly around the world. The World Health Organization (WHO) report reveals, globally the occurrence of obesity is approximately 1.6 billion in adults and provides serious consequences such as 35.8 million people effect with disabilities and 2.8 million is morbidity annually [4,5]. The prevalence rate of obesity is on uttermost in Anglosphere countries inclusive USA, and Middle East high-income countries, which encompass severely obese patients [6,7]. Particularly, in Saudi Arabia the prevalence rate of obesity has been estimated to be 35.4% which is higher than its other neighboring countries 31.7% in United Arab Emirates, and 27% in Oman respectively [8]. A sedentary lifestyle is considered as one of its main reasons across the globe which is estimated at approximately 10% of the population is obese and approximately twice and thrice of 1970s data the reason behind this intake of the calories has increased across the world [9].

Need of awareness of Obesity

Awareness of obesity is mandatory and its early care too. Due to its ignorance, BMI increases rapidly in society. As a result, it provides a severe adverse effect on human health and effecting with several chronic diseases [10], including high blood cholesterol, diabetes, cardiovascular diseases, certain cancers (e.g., Prostrate, breast), sleep apnea and snoring, premature death and osteoarthritis and joint disease. Additionally, it will be a burden to the government too. If we look just 4-5 decades back, this disease was less common and its related chronic diseases too. But now this has been very common and frequently increasing, which results in the intense pressure on other medical and research departments in the hospital such as Cardiology, Endocrinology, Nephrology, Osteopathy, and in some cases of Oncology [11].

Measurement of Obesity

Body mass index (BMI) is a primary method to measure obesity, it has been extensively used by health care professionals for screening, diagnosis, and classification of underweight, overweight, and obese individuals [12]. Globally accepted WHO standards [13] of BMI classification has been shown in Table 1, and its tree structure is presented in Figure 1. It is a sensible and high-quality indicator for the evaluation of obesity, which is on hand, inexpensive and non-invasive for general public [14]. Initially, the BMI equation was developed by Adolphe Quetelet in 1900 [15]. It is defined as the ratio of body mass (in Kg) to the square of body height (in Mt.), and the unit is of measurement is kg/m^2.
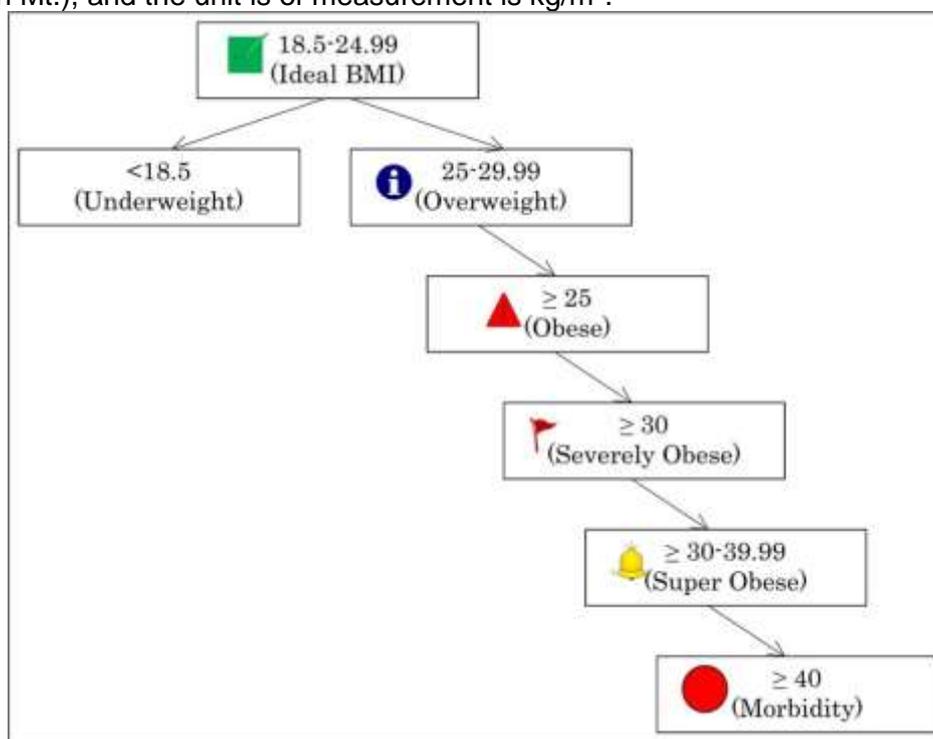

Figure 1. Tree Structure of BMI Classification.

Applications of Data Mining on Chronic Diseases Analysis and Prognosis

Data has been gathered by different sources such as medical data set and biological data sets, but its analysis is one of the most important tasks in order to extract sensible knowledge. For this purpose, data scientists used to take various technological steps based on data size and stakeholder requirements.

Typically, data mining deals with finding relevant logical patterns from the raw data sets based on that discovering knowledge [16]. Its main objective is to retrieve large amounts of meaningful information from unstructured data in the form of structured data, which would be hard to analyze on a particular record basis [17]. The aim of data analysis is to answer accurately, clinical questions with adequate scientific and statistical support. Data mining techniques have been successfully applied in various real-world scenarios - clinical data analysis is one of them. In recent years, applications of data mining have been widely seen on the prediction of the best mode for the intervention of chronic diseases such as hypertension and diabetes. For this, the Oracle Data Miner (ODM) tool has been prominently used in order to explore the meaningful patterns for predicting the effective mode of intervention [18-21]. Few of the related works have been discussed here. Predictive analysis has been done using ODM on diabetes data set to explore the patterns about which intervention type is more effective to which age group (young or old) [18]. Further, to reach some conclusion, research was carried out to find meaningful patterns among two non-communicable diseases: hypertension and diabetes using the ODM tool, it provides very interesting patters that both diseases have a strong relationship like twin sisters and diet management is an important risk factor among them [19]. Different risk factors have been analyzed for hypertension interventions by Naïve Bayesian classifier and found quitting smoking is more effective [20]. A research was carried out to predict which type of treatment is more effective in hypertension using ODM and found that quitting smoking is an effective mode to prevent hypertension [21].

Several researchers have put their efforts into the study of obesity using data mining and machine learning classifiers. SAS and SUDAAN tools had been applied to find the frequency of obesity among American adults and children, it is noted that the percentage of obesity has been increased by 17% [22]. Random forest, SVM and gradient boosting classifiers have been applied on diet data set to predict obesity. Due to the consumption of a high caloric diet, obesity cases increases [23]. Six different types of classifiers were imposed, and logistic regression found more effective for detecting obesity [24]. A survey has been done for predicting obesity using machine learning classifiers, and the performance has been calculated using three parameters - sensitivity, specificity, and ROC. However, the study did not conclude which classifier is best [25]. The Naive Bayesian and decision tree have been used to predictive the obesity, the increase of 20% is noted via Naïve Bayesian (NB) Tree [26]. The prediction has been done using the hybrid approach of data mining classifiers. The outcome shows that the percentage of predictive accuracy is up by 60% to 95% [27]. Artificial Neural Network (ANN) and logistic regression have been used and found that weight reduction plays a vital role in controlling obesity [28]. The traversal study carried out among the female students of Saudi Arabian universities, it concludes that obesity is prevalent and the need for a healthy lifestyle assessment is required [30]. SVM and ANN are applied to breast cancer data set, and they found ANN performs better [29].

Obesity is a major health concern in Saudi Arabia, and is reported proliferated due to the various factors such as consumption of fast food, no sweating activity and unhealthy lifestyle. To the best of our knowledge, similar predictive analysis using AUC of ROC for each BMI category has not been done. Therefore, this research paper explains the predictive analysis using data mining classifier for each BMI category as shown in Figure 1 and Table 1, and analyzes the strong associations of obesity with other chronic diseases. The significant issue for bariatric physicians and overweight/obese patients, is that, this provides the high risk of associated prolonged and chronic diseases such as hypertension, diabetes, cardiovascular disease (CVD), prostrate, arthritis, and some cancers too.

The structure of the paper is described as follows. Section 2 presents the methodology; it includes obesity data collection, description, and database design, applied data mining tools and techniques, and performance evaluation. Section 3, describes the experimental results based on that discussion has been done on each BMI category. Section 4 concludes the paper.

APPLIED METHOD

In this section, we present our applied methodology for predicting the obesity based on BMI values from the World Health Organization (WHO) data set [31] as shown in Figure 2. The data analysis has been performed using Oracle data miner 11g and the outputs are stored in the oracle 11g database server vice versa [32]. The flow of the applied approach has been illustrated in Figure 2. It comprises six steps: data set collection, database design, loading data set, connection to data mining tool, classification, and knowledge discovery.
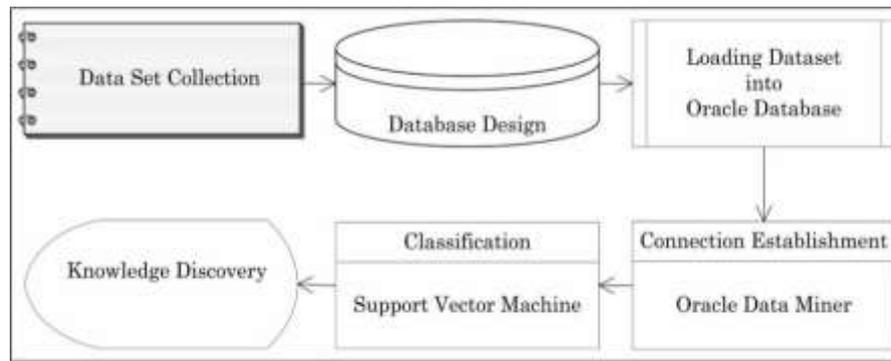
Figure 2. Model for Analyzing the Trends of BMI

Data Set Collection

Data set is a prerequisite for data analysis. In the present research work, we have used the data set of 'Standard Report of Non-Communicable Diseases' and taken obese patients data set, which has been collected from the World Health Organization (WHO) [31]. Based on WHO classification [13], BMI has been classified into seven segments on the basis of distinct BMI values as shown in Table 1.

Table 1. BMI Definition and its Classification.

Classification	BMI Definition (Kg/m^2)
Underweight	< 18.5
Ideal BMI	18.5-24.99
Overweight	25-29.99
Obese	≥ 25
Severely Obese	≥ 30
Super Obese	30-39.99
Morbidity Obese	≥ 40

Database Design

In this work, the database has been designed in Oracle 11g using DBCA (Oracle Database Configuration Assistant) utility. A table segment stores that data for a table that is neither clustered nor partitioned. All the data in a table segment must be stored in one place called tablespace. The designed table includes the set of non-class attributes (predictor) and class (target) attributes. The non-class attributes such as sr_no, year, and sex. The class attribute is 'definition', it has seven distinct class values. The detailed description of the data set has been presented in Table 2.

Table 2. BMI Definition and its Classification.

Attribute	Explanation	Data Type	Constraint
Sr_No	Unique Sequence number to each record	number	Primary key
Year	Year of the survey conducted	date	Not null
Definition	Seven possible class values e.g, BMI ≥ 25	varchar	Not null
Sex	Patient gender, i.e., male/female	varchar	Not null
Age_group	Generalized age of patient, e.g., 15-24	varchar	Not null
Sample_size	Total number of patients in each age group	number	Not null
Prevalence	Disease occurrence to the number of units at risk in the population	number	Not null

Loading Data Set

The primary data set has been loaded into the database schema using SQL*Loader tool for data modelling and predictive analytics. The two-dimensional table comprises total number of 7 attributes and 65 records. For better understanding to readers, we have presented a sample record of applied loaded data set in Table 3. It explains each attribute and its corresponding value. Just to clarify, in each age group, the sample size is static but the prevalence varies because of different BMI values.

Table 3. A sample record of Data Set

Sr_No	Year	Definition	Sex	Age_group	Sample_size	Prevalence
1	2005	BMI ≥ 30	Male	45-54	519	38.9%

Connection Establishment

The Connection establishment is an important task within the process. The model is built by establishing the connection between the ODM and Oracle databases. Primarily the data set is stored in Oracle database. For executing the data mining classifier, the ODM should be connected to the Oracle database. It needs the special database privileges to connect with oracle database server [32].

SQL>GRANT EXECUTE ON CTXSYS.CTX_DDL TO obesity_user.

After executing this Structured Query Language (SQL) query, the ODM get successfully connected to oracle 11g database server. To perform the classification based predictive analysis.

Classification

Classification is a supervised learning method of data mining/machine learning and it plays a significant role in discovering meaningful patterns from the data sets of different domains [17]. The data set has a set of non-class attributes $A_1, A_2, ... A_n$ and a pre-defined class attribute C. It discovers the relationship between non-class attribute and class attribute (target), which accurately predicts the target attribute for each case in the data set. It is used when the class attribute is discrete in nature [34,35]. In the present data set, attribute 'Definition' is the class attribute, and it comprises seven distinct class values of BMI such as <18.5, 18.5-24.99, ≥ 25, 25-29.99, ≥30, 30-39.99, and ≥ 40. Each class value has its own predictive result, and the rest of the attributes are non-class attributes or predictors. In this research work, we have applied supervised learning because the class attribute is defined, and the classifier Support Vector Machine.

Supervised Learning Classifier -- Support Vector Machine (SVM)

The predictive result can be achieved by different types of available machine learning classifiers. It provides a high rank of classification accuracy when the data set have less number of attributes and linearly separates the data into two classes - class1 and class2 [37,38]. Therefore, it has been used for the predictive analysis of different BMI factors. It is based on following Equation 1.

$$f(x, y) = sign(x.a + y), \qquad (1)$$

Where 'x' is a weight vector, 'a' is the input vector and 'y' is the initial point. In ODM, SVM uses binary kernels one is Linear and another is Gaussian based. In our present work, the model building is made of the Gaussian kernel. Applications of SVM have been significantly seen in medical field such as disease diagnosis, prevention, and intervention due to its good predictive results [37]. The study shows, SVM is a suitable classifier for detection and prediction. However, here we used ODM tool because it has the ability to take character variables 'char' as a target attribute as well, it internally processes them into numerical values with the notations as (0,1,2,..) for further processing. In ODM, SVM is a successful classifier for both regression and classification techniques and have strong relationship with oracle database platform which maintains the good ranking of accuracy [39]. After model building, evaluating the performance of each model is one of the most significant tasks. Based on this we can claim that how good is our model? For this, we need to know some basic evaluation parameters such as ROC, AUC and predictive accuracy. To test the performance of each model, ROC curves have been depicted. The better the curve value AUC, better the detection of BMI value.

What is Receiver Operating characteristics (ROC) and its significance in Human Health Research?

ROC is a 2-dimensional graph, between true positive rate (TPR) and false negative rate (FPR), or in other words, it can be defined as the trade-off between sensitivity and specificity [39]. ROC curves play a vital role in clinical data analysis [40,41]. Medical experts use the ROC curve analysis for exploring the performances of diagnosis [42,43]. The measurement of ROC plot is done by Area under curve (AUC).

Area Under Curve (AUC)

The performance of each classified BMI value has been evaluated by the ROC plots. This has been calculated from the area of ROC curve, termed as Area under the curve (AUC). The value of the curve is between 0 and 1. This shows the overall accuracy of the model, if the value reaches 1.0, it is a sign of high sensitivity and specificity [45]. Here, The AUC is used to measure to evaluate the accuracy within each BMI category. An AUC of 1.0 indicates prediction without error of BMI results such as underweight, overweight and obesity status in the data set. On the contrary, an AUC of 0.50 represents a fifty percent possibility of a correct prediction of BMI classifications. The better classifier should be positioned near to the left corner of the height of the ROC plot as highlighted in Figure 3 to 9.

Typically, in medical diagnosis high AUCs are required. As a result, AUC computation has been often applied in neurosciences and endocrinological disorders such as diabetes [45]. In clinical data analysis research, it has been applied to find the patterns and discover which type of treatment is more effective to the patient. For example, quitting smoking is more beneficial to diabetic patients or not. Its computation allows the data scientist to simplify the data analysis with predictive knowledge on health informatics research. The trapezoidal method is used to calculate the AUC. It works by dividing the area into a number of sections of equal width. The summation of the area of each section by the area of the trapezium formed when the upper end is replaced by a chord called the AUC of ROC [33,48,49], which is represented by the trapezoidal rule in Equation 2.

$$T(a, b, n) = \left(\frac{b-a}{n}\right) \times \left(\frac{f(a)-f(b)}{2}\right) + \left(\frac{f[a+i(b-a)]}{n}\right), \quad (2)$$

Furthermore, to verify our results, the predictive accuracy of the models has also been estimated by Equation 3. It has been designed from four parameters TP, TN, FP, and FN as shown in Table 4 with the explanation.

$$Predictive\ Accuracy = \frac{TP+TN}{TP+FP+TN+FN}, \quad (3)$$

Table 4. Prediction types in Obesity case.

Acronym	Prediction Types	Explanation
TP	True Positive	If a patient is 'obese' and also correctly predicted and predicted as 'obese'. For example, BMI ≥ 30 and the classifier also had predicted them as 'obese'.
TN	True Negative	If a person is 'non-obese', and predicted as 'non-obese'. No action is needed.
FP	False Positive	Incorrect prediction, when a healthy person as an 'obese' patient. Then, it creates financial burden.
FN	False Negative	Incorrect prediction, when the classifier predicts the 'obese' patient as a normal person. It is a severe problem in health informatics research.

Knowledge Discovery

Knowledge discovery is an integral part of data mining process. It is the process of extracting relevant and sensible patterns from data sets by combining data mining and statistical machine learning with databases [48,49]. Here the knowledge discovery means, if the SVM predicts about riskier BMI factor, then pattern reveals to other chronic diseases in case of overweight, and for underweight it could be malnutrition and mental illness.

EXPERIMENTAL RESULTS AND DISCUSSIONS

This section deals with the experiment and exploring the trends for each BMI category. The experiment has been done on the obesity data set ***D***, which has been processed using ODM, for analysis we execute the SQL queries and simultaneously the results are saved in log files as database objects called tables in Oracle database. We summarize our results of ROC reported from Figure 4 to 9, and predictive accuracy in Table 5. The trends of each BMI category has been discussed, based on AUC of ROC in subsequent subsections.

Trends obtained from BMI < 18.5

As per [13], if the BMI is less than 18.5, the person is considered as underweight. BMI has a potential relationship with underweight and declares as malnutrition, which results in severe health issues such as mental illness, health-related behaviors, and other biological risk factors [52,53]. Our results as shown in Table 5 and Figure 3, we found that the value AUC is 0.75 which is greater than 0.5, this indicates a good sign of predicting the underweight cases with 90.76% of predictive accuracy. The WHO data set reveals that the prevalence of underweight from the age group 15-64 is less i.e., 4.9% in Saudi Arabia [31]. Despite this low prevalence in the country due to the good economy, still, some cases have been found, it is a sensitive issue. Also, our predictive results show that it is not a good sign, hence, it is suggested that the health department should take necessary medical actions.
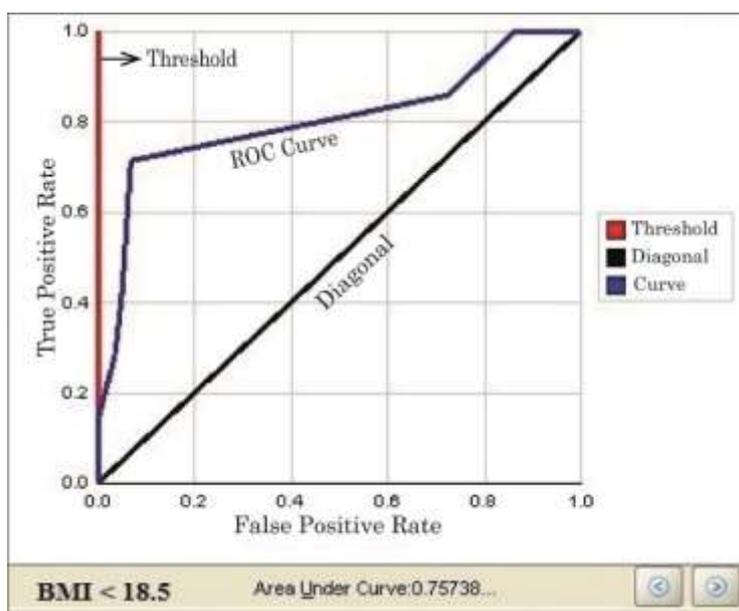

Figure 3. ROC graph for BMI < 18.5

Trends obtained from BMI 18.5-24.99

BMI 18-24.99 is considered as an ideal weight for a healthy person [1]. From primary generalized data set, the prevalence of this classification is found 27.8% which indicates people of the country have to be more focused on their health [30]. Further, we have analyzed from our results as shown in Table 5 and plotted ROC graph in Figure 4, we found a good ranking of the model with a value of AUC is 0.91 and predictive accuracy is 95.38%.
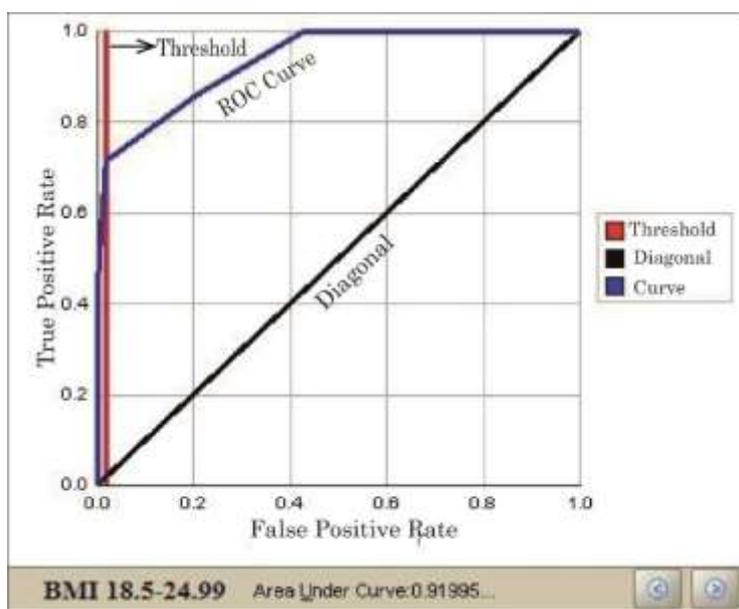

Figure 4. ROC graph for BMI 18.5-24.99

Here, after careful analysis, we can say that this BMI stage of individual is important, and a person has to care about his health in two aspects. Firstly, if the person follows the sedentary lifestyle and takes an unhealthy diet without exercise then the person's BMI may also increase i.e., greater than 25, which results in overweight and this could be an alarming cause for obesity [52]. Secondly, if the person does not take a nutritious healthy diet, a person may suffer from underweight due to the decline in BMI [53].

Trends obtained from BMI 25-29.99

As per WHO criteria, BMI 25-29.99 states that the person is now acknowledged as overweight but not obese [12,53]. The prevalence rate of overweight is 30.73% in the applied data set. However, after predictive analysis, we found the results for this BMI category as shown in Table 5 and plotted ROC graph in Figure 5, shows that the AUC of ROC is 0.93 and its predictive accuracy is 92.30%. Both results show that model is approximately accurate, and the increase of BMI value would result in prevalence of overweight [55]. As a result, overweight person may enter into obesity criteria. We have analyzed from our results and primary data set, prevalence of overweight is higher for age group of 55-64 i.e., 46.4 compare to other age groups. This is a crucial age because person is 60+ (old age) and needs to care for health seriously [56]. Otherwise, this inclination towards the increase in BMI could cause other relevant diseases too.
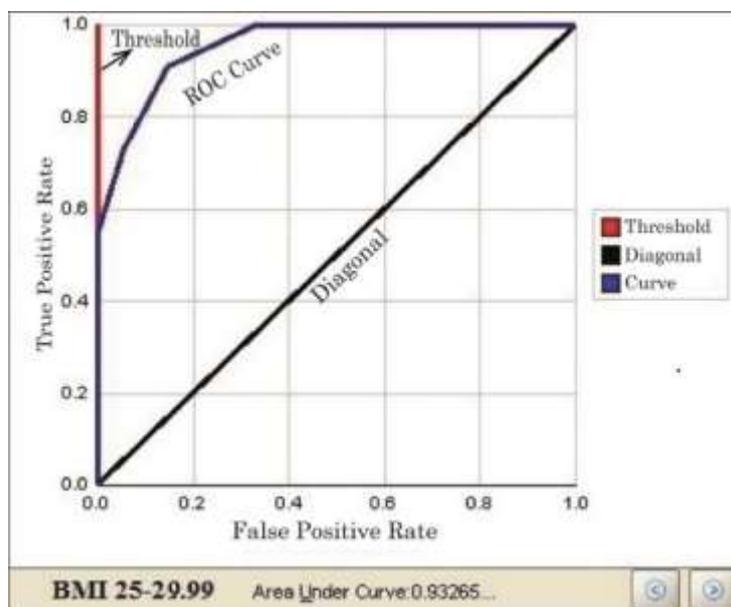

Figure 5. ROC graph for BMI 25-29.99

Trends obtained from BMI ≥ 25

In this BMI category, we will refer person as a patient because BMI ≥ 25 and is said to be in the initial stage of obesity [13,57]. As per primary data set, overall prevalence of obesity for this is 66.2% from the age group 15-64. However, for the age group 45-54 the prevalence rate is 77.8%, it is a perceiving issue. Now, from our predictive results, from Figure 6, the AUC of ROC is 0.99, and the predictive accuracy is 98.46. These results have been presented in Table 5. After a careful review, the results demonstrate that patients with this BMI category enter into the pre-obese stage which is a perceiving issue. Data analytic point of view, our predictive results are appreciable because of the AUC value is near to 1.0 and predictive accuracy is also above 98%, which indicates that the prevalence of patients consisting of this BMI could be approximate double of applied data set in near future and push the patients into higher BMI values. However, patients are at pre-obese stage, but patients must care about diet intake in terms of taking hypocaloric diets, weight reduction and exercise [58]. This will prevent the patient from various related chronic and vascular diseases
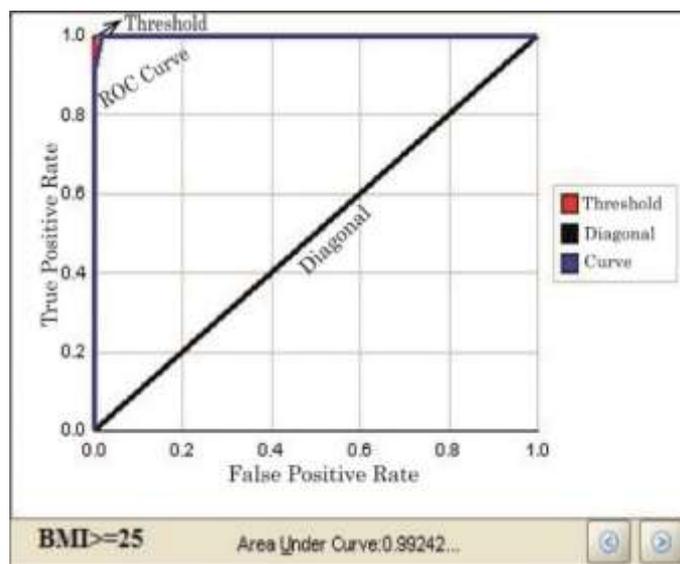
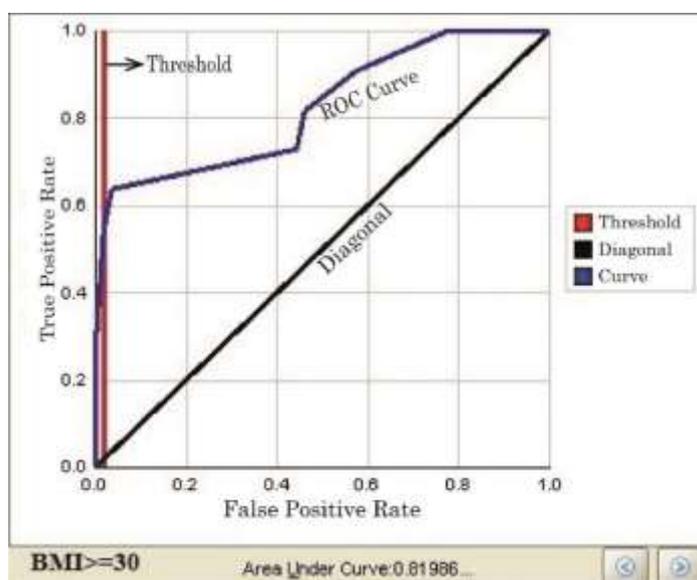**Figure 6.** ROC graph for BMI ≥ 25

Trends obtained from BMI ≥ 30

Figure 7. ROC graph for BMI ≥ 30

When BMI reaches to 30 or greater, then the person is detected as Obese [55,57]. As reported in the applied data set, the prevalence rate for this BMI is 28.3%. Although, it is less than BMI ≥ 25, it is a serious issue for obese patients and bariatric physicians. Here, the concern is that its prevalence is 34.5 in the age group of 35-44, which is higher than other age groups, and this age group is not treated as old age. The AUC of ROC shown in Figure 7 is 0.81 and predictive accuracy is 90.76% as shown in Table 5.

Since 1975, the prevalence of BMI ≥ 30 has been increased in both developed and developing countries [6]. It has been analyzed, with the increase of BMI 30, obesity patients facing Obstructive Sleep Apnea (OSA), and its prevalence rises progressively [52,59]. This happens because of various factors such as ignorance at overweight and pre-obese at the initial stage, obesogenic factors, sedentary lifestyle, and unhealthy diet [55]. The primary data set and predictive results, clearly indicate that the person should take immediate steps to control his BMI corresponding to weight control and other suggested plan by the bariatric physicians. To prevent and treat the related severe chronic health issues [60].

Trends obtained from BMI ≥ 30-39.99

BMI 30-39.99 means the patient is suffering with super-obese of Stage 2 Obesity [1,54]. Primary data set reveals, its overall prevalence in KSA is 25.6%, and 31.7% for the age group 35-44 which higher than other age groups. If we see this previous analysis of BMI ≥ 30, in that too, the prevalence rate was higher for 35-44. This indicates that the influence of risk factor such as unhealthy diet, no weight reduction, lack of

exercise, and obesogenic factors are more to this age group. As a result, increase in BMI. The AUC value is 0.87 and predictive accuracy is 93.84% as shown in Figure 8 and Table 5.
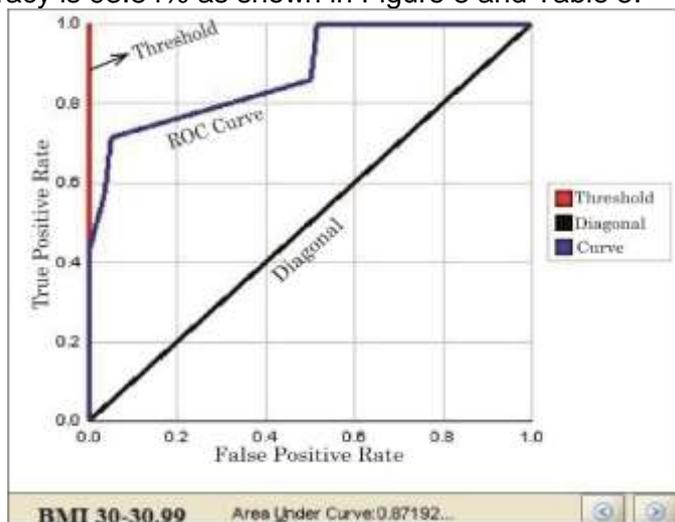

Figure 8. ROC graph for BMI 30-39.99

Our predictive results are also indicating that it will increase in near future too, as the country is suffering from major factors such as a cultural shift towards developed-world, adapting junk diet items regularly, eating habits and lack of physical activity [56,61]. Reduction and control of BMI are significant in this obesity category because patients are more likely to birth other associated chronic diseases. Therefore, it is strongly recommended, the patient has to take regular consultation with the dietitian and bariatric physician.

Trends obtained from BMI ≥ 40

As unnecessary fat increases in our body, BMI increases simultaneously, if it reaches to ≥40 eventually, it is a sensitive issue. As it is considered as a morbidity obese with several acute chronic illnesses [62]. It is estimated in 2015, high BMI majorly affected the global death rate which is approximately around four million [52]. In the plotted ROC plot in Figure 9, the AUC is 0.80, and predictive accuracy is 92.30% as shown in Table 5, which predicts that it may increase the morbidly obese cases in coming future. It is also noted that 63-64% of patients are hypertensive due to BMI≥40 [63,67]. Previously, it was reported obesity is more in developed countries but now this has been drastically increasing in developing countries [7]. Since the last five decades, the prevalence of obesity has been increased rapidly, because of quality life with many causes such as more use of air conditioners, unhealthy lifestyle, less physical activity, addicted to smartphones, less sleep and many more. Which results in the imbalance of calories [65]. Middle East countries have the highest prevalence rates of obesity in adult populations particularly in Saudi Arabia, and it is predicted this will also increase [68]. For proliferated morbidity and mortality, obesity plays a major risk factor, most saliently from cardiovascular disease and diabetes, likewise from cancer and chronic diseases, including osteoporosis, liver and kidney diseases, sleep apnea and depression. Moreover, the availability of junk food items in offices and school canteens, lack of physical activities, and exercises all are related to increase in BMI.

Overall, the data analytics perspective, all seven BMI categories comes in the good rank of data modeling because none of them have an AUC value below 0.5. The analysis reveals interesting results, as tabulated in Table 5, the AUC value and predictive accuracy for BMI ≥ 25 is high which are approximately accurate in comparison to other BMI definitions and the important thing is that this stage is pre-obese. If it comes in the eye to both patient and bariatric physician, they have to carefully take the required necessary actions in order to prevent this disease and its related chronic diseases. If it cannot intervene at this BMI category, it has been predicted this will enhances linearly and crosses the edge that has been observed its direct relationship to other related diseases due to the increase of BMI, the risk is also higher for numerous diseases such as cardiovascular, high blood pressure, diabetes, prostate, osteoarthritis, and some cancers (e.g., breast, and gallbladder) [66]. Most of the associated disease results in prolonged illness except in some cases of cancer.
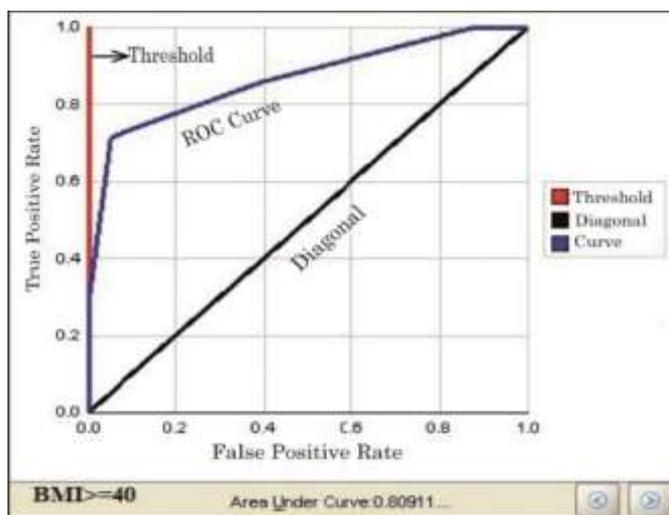
Figure 9. ROC graph for BMI ≥ 40

Table 5. Predictive results of BMI Category with AUC and predictive accuracy

Classification	BMI Category	AUC	Predictive Accuracy (%)
Underweight	<18.5	0.75	90.76
Ideal BMI	18.5-24.99	0.91	95.38
Overweight	25-29.99	0.93	92.30
Obese	**≥ 25**	**0.99**	**98.46**
Severely Obese	≥ 30	0.81	90.76
Super Obese	30-39.99	0.87	93.84
Morbidly Obese	≥ 40	0.80	92.30

Furthermore, it can lead to the chance of morbidly obese. As a result, obesity is a significantly contributes in morbidity and mortality rates [68]. For this, public awareness at primary level is mandatory like other diseases [70,71,72] for the interventions to prevent the environment from overweight or obesity. Secondary, if the overweight is found promote the weight loss methods in society, excess exercise and aware them to be away from sedentary lifestyle and proper prescribed lifestyle management should be adopted. In order to prevent them from serious complications. Last, both prevention and treatment are required by bariatric physicians to control the related health complications. Hence, our predictive results will provide aid to therapeutic decisions. Moreover, We have also analyzed BMI < 18.5 has a strong relationship with underweight and person can be declared as malnutrition, which results in severe health issues such as mental illness, health-related behaviors, and other biological risk factors.

CONCLUSION

The key contribution of this research work is the idea of predicting each BMI category from underweight to morbid obese using the classifier SVM and the concept of AUC of ROC with predictive accuracy. The predictive analysis has been done on seven distinct types of BMI categories of the data set. The main issue here is the interpretation of data analysis as per the health care unit. So that it will assist human beings and bariatric physicians. As shown in Table 5, each BMI results have the AUC value greater than 0.75, and a Predictive accuracy is above 80%. However, BMI ≥ 25 has the highest AUC of ROC 0.99, and its predictive accuracy is 98.46 respectively, which shows that the predictive model is approximately accurate. After a careful review of our results and relating with medical literature, we came to the conclusion that BMI ≥ 25 is an alarming indication and if not prevented on time as directed by a physician, it will reflect the sign of other prolonged chronic diseases due to the gradual increase in BMI. This research also gives the awareness to the society and its further severe health consequences.

Conflicts of Interest: The authors declare that they have no conflict of interest.

REFERENCES

1. Organization WH. Obesity: preventing and managing the global epidemic report of a WHOconsultation (WHO technical report series 894); Available from: https://www.who.int/nutrition/publications/obesity/WHO_TRS_894/en/ [Cited on 2019 Sep]
2. Roser HRaM. Obesity. Available online: https://ourworldindata.org/obesity (Accessed on Oct 2019).
3. Rossen L, Rossen E. Obesity 101. Springer Publishing Company; 2011.
4. Delgado-López PD, Castilla-Díez JM. Impact of obesity in the pathophysiology of degenerative disk disease and in the morbidity and outcome of lumbar spine surgery. Neurocirugía (English Edition) 2018;29(2):93-102.
5. WHO. WHO Global Health Observatory (GHO) data. Overweight and Obesity. 2014; Available online: https://www.who.int/gho/ncd/risk_factors/overweight/en/## [Cited on 2019 Dec]
6. Collaboration NRF. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19· 2 million participants. The Lancet 2016;387(10026):1377-96.
7. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. The lancet 2014;384(9945):766-81.
8. Al-Raddadi R, Bahijri SM, Jambi HA, Ferns G, Tuomilehto J. The prevalence of obesity and overweight, associated demographic and lifestyle factors, and health status in the adult population of Jeddah, Saudi Arabia. Therapeutic Advances in Chronic Disease 2019;10:2040622319878997.
9. Nguyen DM, El-Serag HB. The epidemiology of obesity. Gastroenterology Clinics 2010;39(1):1-7.
10. Pi-Sunyer X. The medical risks of obesity. Postgraduate medicine 2009;121(6):21-33.
11. Park H-K, Ahima RS. Endocrine Disorders Associated with Obesity. In: Ahima RS, editor. Metabolic Syndrome: A Comprehensive Textbook. Cham: Springer International Publishing; 2016. p. 743-59.
12. Garvey WT. Clinical Definition of Overweight and Obesity. In: Bariatric Endocrinology: Springer; 2019. p. 121-43.
13. Organization WH. Obesity: preventing and managing the global epidemic. World Health Organization; 2000.
14. Suchanek P, Kralova Lesna I, Mengerova O, Mrazkova J, Lanska V, Stavek P. Which index best correlates with body fat mass: BAI, BMI, waist or WHR. Neuro Endocrinol Lett 2012;33(Suppl 2):78-82.
15. Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. Nutrition today 2015;50(3):117.
16. Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science 2001;16(3):199-231.
17. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier; 2011.
18. Aljumah AA, Ahamad MG, Siddiqui MK. Application of data mining: Diabetes health care in young and old patients. Journal of King Saud University-Computer and Information Sciences 2013;25(2):127-36.
19. Aljumah A, Siddiqui M. Data Mining Perspective: Prognosis of Life Style on Hypertension and Diabetes. International Arab Journal of Information Technology (IAJIT) 2016;13(1):93-9.
20. Aljumah AA, Siddiqui MK. Hypertension interventions using classification based data mining. Research Journal of Applied Sciences, Engineering and Technology 2014;7(17):3593-602.
21. Almazyad AS, Ahamad MG, Siddiqui MK, Almazyad AS. Effective hypertensive treatment using data mining in Saudi Arabia. Journal of clinical monitoring and computing 2010;24(6):391-401.
22. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. Jama 2014;311(8):806-14.
23. Dunstan J, Aguirre M, Bastías M, Nau C, Glass TA, Tobar F. Predicting nationwide obesity from food sales using machine learning. Health informatics journal 2019:1460458219845959.
24. Zhang S, Tjortjis C, Zeng X, Qiao H, Buchan I, Keane J. Comparing data mining methods with logistic regression in childhood obesity prediction. Information Systems Frontiers 2009;11(4):449-60.
25. Butler ÉM, Derraik JG, Taylor RW, Cutfield WS. Prediction Models for Early Childhood Obesity: Applicability and Existing Issues. Hormone research in paediatrics 2018;90(6):358-67.
26. Adnan M, Husain W, Rashid N. IMPLEMENTATION OF HYBRID NAIVE BAYESIAN-DESICION TREE FOR CHILDHOOD OBESITY PREDICTIONS. network 2018; 40:84-7.
27. Adnan MHM, Husain W. Hybrid approaches using decision tree, naive Bayes, means and euclidean distances for childhood obesity prediction. International Journal of Software Engineering and Its Applications 2012;6(3):99-106.
28. Lee Y-C, Lee W-J, Lee T-S, Lin Y-C, Wang W, Liew P-L, et al. Prediction of successful weight reduction after bariatric surgery by data mining technologies. Obesity surgery 2007;17(9):1235-41.
29. Al NQ. Obesity among Saudi Female University Students: Dietary Habits and Health Behaviors. The Journal of the Egyptian Public Health Association 2010;85(1-2):45-59.
30. Al-Hamdan N, Kutbi A, Choudhry A, Nooh R, Shoukri M, Mujib S. WHO stepwise approach to NCD surveillance country-specific standard report Saudi Arabia. 2005; Available from: https://apps.who.int/infobase/Indicators.aspx

31. Corporation O. Oracle Data Miner. Oracle Corporation; Available from: https://www.oracle.com/database/technologies/datawarehouse-bigdata/dataminer.html [Cited on 2018 Sep]

32. Tamayo P, Berger C, Campos M, Yarmus J, Milenova B, Mozes A, et al. Oracle data mining. In: Data mining and knowledge discovery handbook: Springer; 2005. p. 1315-29.

33. Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Singh V. Application of Deep Learning for Fast Detection of COVID-19 in X-Rays using nCOVnet. Chaos, Solitons & Fractals. 2020 May;138:109944.

34. Siddiqui MK, Islam MZ, Kabir MA. Analyzing Performance of Classification Techniques in Detecting Epileptic Seizure. In: International Conference on Advanced Data Mining and Applications: Springer; 2017. p. 386-98.

35. Vapnik V. The nature of statistical learning theory. Springer science & business media; 2013.

36. Rose RR, Meena K, Suruliandi A. An Empirical Evaluation of the Local Texture Description Framework-Based Modified Local Directional Number Pattern with Various Classifiers for Face Recognition. Brazilian Archives of Biology and Technology 2016;59(SPE2).

37. Bentaouza C, Benyettou M, Abe S, Alphonso I, Bishop C, Boser B, et al. Support Vector Machines: Theory and Application. Journal of Applied Sciences 2005;10(16):pp: 144-52.

38. Milenova BL, Yarmus JS, Campos MM. SVM in oracle database 10g: removing the barriers to widespread adoption of support vector machines. In: Proceedings of the 31st international conference on Very large data bases: VLDB Endowment; 2005. p. 1152-63.

39. Streiner DL, Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. The Canadian Journal of Psychiatry 2007;52(2):121-8.

40. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Physics in Medicine & Biology 2018;63(7):07TR1.

41. Wen P, Chen S, Wang J, Che W. Receiver Operating Characteristics (ROC) analysis for decreased disease risk and elevated treatment response to pegylated-interferon in chronic hepatitis B patients. Future Generation Computer Systems 2019;98:372-6.

42. Hart PD. Receiver operating characteristic (ROC) curve analysis: A tutorial using body mass index (BMI) as a measure of obesity. J Phys Act Res 2016;1:5-8.

43. Pérez-Ruixo C, Remmerie B, Peréz-Ruixo JJ, Vermeulen A. A Receiver Operating Characteristic Framework for Non-adherence Detection Using Drug Concentration Thresholds—Application to Simulated Risperidone Data in Schizophrenic Patients. The AAPS journal 2019;21(3):40.

44. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain 2008;8(6):221-3.

45. Huang C-C, Chung C-M, Leu H-B, Lin L-Y, Chiu C-C, Hsu C-Y, et al. Diabetes mellitus and the risk of Alzheimer's disease: a nationwide population-based study. PloS one 2014;9(1):e87095.

46. Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology 2003;28(7):916-31.

47. Yeh S-T. Using trapezoidal rule for the area under a curve calculation. Proceedings of the 27th Annual SAS® User Group International (SUGI'02) 2002.

48. Elloumi M, Zomaya AY. Biological knowledge discovery handbook: Preprocessing, mining and postprocessing of biological data. John Wiley & Sons; 2013 Dec 24.

49. Aguiar-Pulido V, A Seoane J, Gestal M, Dorado J. Exploring patterns of epigenetic information with data mining techniques. Current pharmaceutical design 2013;19(4):779-89.

50. Ng W, Collins P, Hickling DF, Bell JJ. Evaluating the concurrent validity of body mass index (BMI) in the identification of malnutrition in older hospital inpatients. Clinical Nutrition 2019;38(5):2417-22.

51. Lorem GF, Schirmer H, Emaus N. What is the impact of underweight on self-reported health trajectories and mortality rates: a cohort study. Health and quality of life outcomes 2017;15(1):191.

52. Collaborators GO. Health effects of overweight and obesity in 195 countries over 25 years 2017 [13-27].

53. Agarwal E, Ferguson M, Banks M, Vivanti A, Batterham M, Bauer J, et al. Malnutrition, poor food intake, and adverse healthcare outcomes in non-critically ill obese acute care hospital patients. Clinical Nutrition 2019;38(2):759-66.

54. Organization WH. Global Health Observatory (GHO) data: obesity. World Health Organization. Available on: www. who.int/gho/ncd/risk_factors/overweight/en/ (accessed on 11/11/2019).

55. Gubur S. Determination of the Effect of the Elimination Diet Applied for Overweight and Obese People with Food Intolerance on Body Composition and Biochemical Parameters. Brazilian Archives of Biology and Technology 2018;61.

56. Alturki HA, Brookes DS, Davies PS. Comparative evidence of the consumption from fast-food restaurants between normal-weight and obese Saudi schoolchildren. Public health nutrition 2018;21(12):2280-90.

57. Lipsky LM, Haynie DL, Hill C, Nansel TR, Li K, Liu D, et al. Accuracy of Self-Reported Height, Weight, and BMI Over Time in Emerging Adults. American journal of preventive medicine 2019;56(6):860-8.

58. Scheen AJ, Paquot N. Nutritional counseling for overweight patients and patients with metabolic syndrome. In: Cardiovascular Prevention and Rehabilitation: Springer; 2007.p. 201-11.

59. Lloyd L, Langley-Evans S, McMullen S. Childhood obesity and risk of the adult metabolic syndrome: a systematic review. International journal of obesity 2012;36(1):1.

60. Luig T, Elwyn G, Anderson R, Campbell-Scherer DL. Facing obesity: Adapting the collaborative deliberation model to deal with a complex long-term problem. Patient education and counseling 2019;102(2):291-300.

61. Pitanga FJ, Alves CF, Pamponet ML, Medina MG, Aquino R. Combined effect of physical activity and reduction of screen time for overweight prevention in adolescents. Revista Brasileira de Cineantropometria & Desempenho Humano 2019;21.

62. Banez LL, Albisinni S, Freedland SJ, Tubaro A, De Nunzio C. The impact of obesity on the predictive accuracy of PSA in men undergoing prostate biopsy. World journal of urology 2014;32(2):323-8.

63. Silva GECd, Bazotte RB, Curi R, Silva MARCP. Investigation of risk factors to coronary heart disease in two countryside villages. Brazilian Archives of Biology and Technology. 2004;47(3):387-90.

64. Kopelman PG. Obesity as a medical problem. Nature 2000;404(6778):635.

65. Kyrou I, Randeva HS, Tsigos C, Kaltsas G, Weickert MO. Clinical problems caused by obesity. In: Endotext [Internet]: MDText. com, Inc.; 2018.

66. Lavie CJ, Arena R, Alpert MA, Milani RV, Ventura HO. Management of cardiovascular diseases in patients with obesity. Nature Reviews Cardiology 2018;15(1):45.

67. Aljumah AA, Ahamad MG, Siddiqui MK. Predictive analysis on corre treatment using data mining approach in Saudi Arabia. 2011;3(6):252-61

68. Hensrud DD, Klein S. Extreme obesity: a new medical crisis in the United States. InMayo Clinic Proceedings 2006 Oct 1 (Vol. 81, No. 10, p. S5-S10). Elsevier.

69. Aljumah AA, Siddiqui MK, Ahamad MG. Application of Classification based Data Mining Technique in Diabetes Care. May 2013;13(3): 416-22.

70. Siddiqui MK, Morales-Menendez R, Huang X, Hussain N. A review of epileptic seizure detection using machine learning classifiers. Brain Informatics. 2020 Dec;7(1):1-8.

71. Siddiqui MK, Morales-Menendez R, Gupta PK, Iqbal HM, Hussain F, Khatoon K, Ahmad S. Correlation between temperature and COVID-19 (suspected, confirmed and death) cases based on machine learning analysis. J Pure Appl Microbiol. 2020;14(suppl 1):1017-24.

72. Siddiqui MK, Islam MZ, Kabir MA. A novel quick seizure detection and localization through brain data mining on ECoG dataset. Neural Comput & Applic. 2019 Sep 1;31(9):5595-608.