Machine Learning Approach to Study the Impact of Obesity on Autonomic Nervous System using Heart Rate Variability Features

SR Rathod
Department of Instrumentation and Control Engineering
College of Engineering Pune
Pune, India
rathodsr@yahoo.com

UM Chaskar
Department of Instrumentation and Control Engineering
College of Engineering Pune
Pune, India
u_chaskar@yahoo.com

CY Patil
Department of Instrumentation and Control Engineering
College of Engineering Pune
Pune, India
cypatil@gmail.com

2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII) | 978-1-6654-4126-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICBSII51839.2021.9445129

***Abstract*— Obese people have high chances of cardiovascular disease (CVD), which is supposed to be due to the alteration in autonomic nervous system (ANS) activity. The changes in ANS activity can be identified using heart rate variability (HRV). HRV is a non-invasive tool to measure the ANS activity using linear and non-linear HRV features. The paper presents an aim to understand the effect of obesity on ANS using HRV parameters. Initially, sixteen control and sixteen obese subjects of both the gender between ages 20 to 50 were involved in the study after that synthetic minority oversampling technique (SMOTE) was used to increase the sample size of control and obese subjects from sixteen to forty-eight. The statistically significant difference between two groups was observed using the Independent t test. The statistical results of the study indicate the sympathovagal imbalance due to reduced parasympathetic activity. The statistical results were validated by incorporating the machine learning technique into the study. Machine Learning (ML) algorithm helps to identify the most important predictor that can clearly differentiate control and obese subjects. The statistical and ML algorithm result shows changes in the sympathovagal balance due to decreased parasympathetic activity.**

Keywords— Obesity, Cardiovascular disease (CVD), Autonomic nervous system (ANS), Heart rate variability (HRV), Synthetic minority oversampling technique (SMOTE), Machine Learning (ML).

I. Introduction

One of the leading disorder that enhance mortality in an obese person. The definition of obesity says an excessive fat accumulation in the body that resulted in chronic diseases like hypertension, CVD, myocardial infarction (MI), and diabetes. Many researchers have found a strong correlation between obesity and CVD [1]. The study has suggested that an imbalance of autonomic activity increases CVD chances in obesity [2]. The ANS is a control mechanism of the body that generally maintains homeostasis in the body. ANS regulates the glands, blood vessels, and internal organs. The ANS is divided into two branches sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS mobilizes the body systems to provide energy for the fight or flight response, whereas PNS conserve the energy by regulating the rest and digest response.

HRV measures the effect of the ANS function on heart as the vagal nerve is the mediator between ANS and heart. Even a small change in ANS resulted in changes in heart rhythm. HRV is a variation in the RR interval of electrocardiogram (ECG). Thus, HRV could be the most important and non-invasive method to investigate the impact of obesity on ANS. The significantly decreased HRV in obesity increase the chances of CVD[2,3]. ANS control vital organs of the body, fluctuation in these organs can be represented using linear and non-linear HRV parameters.

The paper is organized as follows- Section II presents the methodology where subjects, criteria for obesity, statistical test, and machine learning algorithm are discussed, section III discussed the statistical and machine learning results , and finally, the conclusion is given in section IV.

II. Methodology

A. Subject

This study was performed solely for research purposes at the institute level with the permission of Dean Research and Development of College of Engineering Pune following all ethical guidelines. The researcher and subjects have made a voluntary agreement. The study involves the electrocardiogram(ECG) acquisition of sixteen normotensive obese individuals and sixteen control subjects between 20 to 50 years of age of both genders who participated in the study. However, sixteen sample size of control and obese are not sufficient to analyse the statistical results. Thus we have synthetically increased the sample size of control and obese subjects using the Synthetic Minority Oversampling Technique(SMOTE)[4]. It is powerful and most widely used technique. It creates random set of samples to balance minority class. New synthetic data samples are generated between randomly chosen minority class sample and its nearest neighbors samples. The details about the implementation of SMOTE technique is given in Algorithm 1.

Algorithm 1 SMOTE Algorithm

Input : Dataset $D, \{y_i \epsilon\, T\}$ where $i = 1, 2, \ldots T$

Number of minority samples (T)

SMOTE Percentage (P)

Number of nearest neighbors (k)

Output : Synthetic Samples

for $i = 1, 2, \ldots T$ **do**

Find the k-nearest neighbors of y_i

$\hat{P} = \left[\frac{P}{100}\right]$

while $\hat{P} \neq 0$ **do**

Select randomly k-nearest neighbors $y_{n(i)}$

Choose randomly $\delta\ \epsilon\ [0, 1]$

$y_{new} = y_i + \delta\,(y_{n(i)} - y_i)$

$T \leftarrow T + y_{new}$

$\hat{P} = \hat{P} - 1$

end while

end for

B. Criteria to decide

As per the World health organization(WHO) guidelines, obesity is determined using BMI. The BMI is calculated as BMI = Weight(kg)/Height(m2)[10]. Subjects with a BMI value of 18 to 25(kg/m^2) are considered normal or non-obese, and subjects with a BMI value of more than 30(kg/m^2) are considered obese[5, 6].

C. ECG Recording and HRV Analysis

The ECG of sixteen obese and control subjects was recorded using a standard ECG system at sampling frequency of 500Hz in a supine resting position for 15 minutes. The last five minutes segments were used for HRV analysis in the Heart rate variability analyzer of Biomedical Workbench LabVIEW. The parameters of HRV were determined using a linear and non-linear method[7]. The linear process involves the analysis of the time domain and frequency domain. In the time domain, the RR interval signal is used to extract the statistical parameters such as mean HR, mean RR, SDNN, and RMSSD. The mean HR and mean RR represents the average values of heart rate and RR interval. The SDNN indicates the standard deviation of the normal to normal(NN) interval, and RMSSD represents the root mean square of the standard deviation of the NN interval. The Fast Fourier Transform (FFT) technique was used to calculate the frequency domain parameters of HRV. The frequency-domain parameters that are extracted from HRV are total power(TP)(ms^2), low frequency(LF) and high frequency(HF) power, low frequency(LF) and high frequency(HF) in normalized unit, and LF/HF ratio. The two non-linear HRV features are extracted from the HRV signal, i.e., SD1 and SD2. The SD1 and SD2 are two Poincare Plot measures. SD1 represents the short term variability in the NN interval, and SD2 is the long term variability in NN interval. The control and obese group were analyzed based on the linear and non-linear HRV parameters[8,9,10,11,12].

D. Statistical Analysis

The statistically significant difference between the control and obese group was observed using the Independent t test. The results are presents as mean ± standard deviation. A level of significance $p < 0.05$ was considered statistically significant.

E. Machine Learning Algorithm

While analyzing the control and obese group using a statistical test, it was observed that most of all, the HRV linear and non-linear parameters show a significant difference. But the important predictor that can differentiate the control and obese group was not identified using a statistical test. Thus we have used a non-linear machine-learning algorithm to find important predictors between control and obesity. In this study, we have used two non-linear machine learning algorithms, i.e., Classification and Regression Tree(CART)[13] and Gradient Boosting Decision Tree(GBDT)[14].

Both of these algorithms are used to find important predictors. The important predictor was found out using the feature importance score. Feature importance indicates the importance of each features. A feature importance score greater than 90% is considered an important predictor. The important predictor feature will be given as input to the ML algorithm, and its performance was evaluated using performance metrics.

F. Performance Metrics

Six different classification quality evaluation measurements such as accuracy, sensitivity, specificity, precision, F1 score, and Area under the receiver operating characteristic curve (AUC) were used. These classification measures calculated using the following confusion matrix.

TABLE I. CONFUSION MATRIX

		Actual Value	
		Positive	**Negative**
Predicted Value	**Positive**	TP	FP
	Negative	FN	TN

TP- True Positive, FP- False Positive, FN- False Negative, TN-True negative

The details about the performance measure are as follow-

1. **Accuracy** is the ratio of the total number of instances of the correct prediction. Accuracy calculated as follows-

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{... (1)}$$

2. **Sensitivity** is used to determine the portion of the actual positive instances case classified adequately by the classifier. Sensitivity calculated as follows-

$$Sensitivity = \frac{TP}{TP + FN} \quad \text{... (2)}$$

3. **Specificity** is used to know the ability of classifiers to identify incorrectly classified negative cases.

$$Specificity = \frac{TN}{TN + FP} \quad \text{... (3)}$$

4. **Precision** It is an indicator that defines the true portion of the instances when predicted to be true. Precision calculated as follows-

$$Precision = \frac{TP}{TP + FP} \quad \text{... (4)}$$

5. **F1 Score** is a harmonic mean of recall and precision. It must be one for good performance and zero for the bad performance of the classification algorithm. F1 score calculated as follows-

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad \text{... (5)}$$

6. **Area under ROC curve** is another important metrics to evaluate the performance of machine learning algorithm. AUC near to 1 indicates the perfect performance and near to 0.5 indicates worst performance of the machine learning model [15].

III. Results And Discussion

A. Statistical Results

Time-domain HRV Parameters

The time-domain HRV parameters mean RR, SDNN, and RMSSD were significantly reduced in the obese group. Reduced mean RR and SDNN indicate that the RR interval time series signal variability is reduced, and total variance is also reduced. The lower value of RMSSD represents reduced parasympathetic activity(Table II).

TABLE II. Time Domain HRV Parameters

Time Domain HRV Features	Healthy Group (n=48)	Obese Group (n= 48)	P-Value
mean HR	67.33 ± 9.82	72.10 ± 10.51	0.0237
mean RR	855.17 ± 124.75	798.69 ± 116.47	0.0241
SDNN	57.42 ± 8.49	49.90 ± 7.50	0.0001
RMSSD	47.44 ± 7.25	40.23 ± 6.11	0.0001

Frequency-domain HRV Parameters

In the frequency domain, TP(ms^2), HF(ms^2), HF(nu), and LF: HF ratio were significantly decreased whereas LF(ms^2) and LF(nu) were comparable in obese as compared to control. Reduced values of TP (ms^2) indicate the less variance. A lower value of HF (ms^2) indicates the declined parasympathetic tone. Less value of LF: HF ratio indicates the autonomic imbalance (Table III).

TABLE III. Frequency Domain HRV Parameters

Frequency Domain HRV Features	Healthy Group (n=48)	Obese Group (n= 48)	P-Value
TP (ms^2)	3265.70 ± 537.58	2543.04 ± 451.68	0.0001
LF (ms^2)	948.12 ± 156.55	926.26 ± 171.12	0.5154
HF(ms^2)	1011.06 ± 171.96	637.48 ± 111.17	0.0001
LF (nu)	48.07 ± 7.16	57.11 ± 8.36	0.0001
HF (nu)	49.94 ± 7.42	40.88 ± 6.07	0.0001
LF:HF	1.59 ± 0.26	1.08 ± 0.16	0.0001

Non-linear HRV Parameters

The non-linear HRV parameters SD1 and SD2 were analyzed and found less in obese compared to control. The SD1 feature value was significantly less in obese that indicates the reduced short-term variability in the HRV signal (Table IV).

TABLE IV. Non Linear HRV Parameters

Non-linear HRV Features	Healthy Group (n=48)	Obese Group (n= 48)	P-Value
SD1	33.72 ± 5.07	28.11 ± 4.30	0.0001
SD2	66.00 ± 9.62	65.83 ± 9.70	0.9303

B. Machine Learning Results

The ML algorithms are used to find the most important predictor that separates obese subjects from the control. However, in the statistical analysis, it was observed that most of the time domain, frequency domain, and non-linear HRV parameters are significantly reduced but do not give an important predictor. The important predictor can be found out using the feature importance technique, which provides a feature importance score to each feature. The feature importance score greater than 0.90 or 90%, indicates the most important predictor.

In this study, we found that mean RR, LF: HF ratio and HF (ms^2) was the most important predictor. We have obtained mean RR and LF: HF important predictor using the CART algorithm, whereas HF (ms^2) was the important predictor obtained using the GBDT. We have used only these predictors as input to the CART and GBDT ML algorithm. When we have applied the mean RR and LF: HF ratio as input to the CART algorithm, we got an accuracy of 96.55%, a sensitivity of 100%, a specificity of 92.86%, precision of 93.75%, F1 score of 0.96 with an AUC of 0.96. When we have used HF (ms^2) as input to the GBDT algorithm, we observed accuracy of 93.10%, the sensitivity of 93.33%, the specificity of 92.86%, precision of 93.33%, F1 score of 0.93 with an AUC of 0.92.

The important predictor indicates that the CART and GBDT ML algorithm can classify the obese and control subjects with an accuracy of 96.55% and 93.10%, respectively (Table V).

TABLE V. Evaluation of Important HRV Predictor

Algorithm	Important HRV Predictor	Feature Importance Score	AC (%)	SE (%)	SP (%)	PR (%)	F1 Score	AUC
CART	**mean RR**	0.93	96.55	100	92.86	93.75	0.96	0.96
	LF/HF	0.94						
GBDT	**HF(ms^2)**	0.95	93.10	93.33	92.86	93.09	0.93	0.92

This suggests that obesity modulates the function of ANS, as mean RR, HF (ms^2), and LF: HF ratio are reduced in obese subjects. Thus changes in ANS modulates cardiac activity.

IV. CONCLUSION

In the present study, we have used short-term HRV analysis of obese and control subjects to study the impact of obesity on ANS. We have used real and synthetic HRV data for analysis. The results of the study are presented using statistical tests and a machine learning algorithm. The statistical analysis shows the significant reduction in HRV parameters of obese subjects compared to control and machine learning algorithm was used to found important HRV predictor. The statistical results suggests an alteration in sympathovagal balance due to less parasympathetic activity. Further, this was confirmed using the CART and GBDT algorithm, which showed a classification accuracy of 96.55% and 93.10%, respectively.

REFERENCES

[1] Hurt RT, Kulisek C, Buchanan LA, McClave SA. The obesity epidemic: challenges, health initiatives, and implications for gastroenterologists. Gastroenterol Hepatol (N Y). 2010;6(12):780792.

[2] Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int J Cardiol. 2009; 141(2):122131.

[3] Rossi RC, Vanderlei LC, Gonalves AC, Vanderlei FM, Bernardo AF, Yamada KM, da Silva NT, de Abreu LC. Impact of obesity on autonomic modulation, heart rate and blood pressure in obese young people. Autonomic neuroscience. 2015;193:138-41.

[4] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002 ;16:321-57.

[5] World Health Organization: obesity and overweight. Fact sheet No. 311,[updated June 2016]. Available from: http://www.who.int/mediacentre/factsheets/fs311/en/. Accessed December 13, 2016.

[6] Romero-Corral A, Somers VK, Sierra-Johnson J, et al. Accuracy of bodymass index to diagnose obesity in the US adult population. Int J Obes(Lond). 2008;3(6):959966.

[7] Heart rate variability: standards of measurement, physiological interpretationand clinical use. task force of the European society of cardiology and the North American Society of Pacing and Electrophysiology. Circulation. 1996;93(5):10431065.

[8] Vanderlei LCM, Pastre CM, Hoshi RA, Carvalho TD, Godoy MF. Basic notions of heart rate variability and its clinical applicability. Rev Bras Cir Cardiovasc. 2009;24(2):205217.

[9] Laederach-Hofmann K, Mussgay L, Ruddel H. Autonomic cardiovascular regulation in obesity. J Endocrinol. 2000;164(1):5966.

[10] Billman GE, Huikuri HV, Sacha J, Trimmel K. An introduction to heart rate variability: methodological considerations and clinical applications. Front Physiol. 2015;6:55.

[11] Billman GE. The effect of heart rate on the heart rate variability response to autonomic interventions. Frontiers in physiology. 2013;4:222.

[12] Shenoy AR, Doreswamy V, Shenoy JP, Prakash VS. Impact of obesity on cardiac autonomic functions in middle aged males. National Journal of Physiology, Pharmacy and Pharmacology. 2014;4(3):236-9.

[13] Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. CRC press; 1984.

[14] Natekin A, Knoll A. Gradient boosting machines, a tutorial. Frontiers in neurorobotics. 2013;7:21.

[15] Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian Journal of internal medicine. 2013;4(2):627-635.