

Obstructive Sleep Apnea Detection using Machine Learning Techniques

Mr. Pendli Venkatesh

Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract— Many of the research studies that have focused on the issue of sleep apnea conditions among the people, emphasize the fact that the numbers are rising in significant numbers year on year. Profoundly, identifying symptoms in the patients is very important to ascertain the possible impact of sleep apnea in patients. The researchers in earlier studies have focused on the conditions of systematical physical examination over the patients who are prone to physical examination for head and neck aches, has relative impact of the osa conditions and also on some scoring-based models using the machine learning solutions. The scope of a new model could be about identification of the features in two stage model. The first stage could be about understanding the lifestyle and psychological conditions of the patient data and accordingly choose the metrics and the model of osa detection tool that can be used for analysis and detecting sleep apnea using Logistic Regression, Linear Regression and K-Means clustering in Machine Learning. If such comprehensive approach can be developed, it can be effective process for developing a sustainable solution.

I. INTRODUCTION

Obstructive sleep apnea (OSA) is the most common type of sleep apnea and is caused by complete or partial obstructions of the upper airway. Recent studies suggest that about 10% of the population is suffering from OSA. Most cases of OSA are believed to be caused by Old age, decreased muscle tone, brain injury (temporary or permanent), Structural features that give rise to a narrowed airway and Increased soft tissue around the airway. Numerous treatment options are used in OSA. One of the traditional methods to treat OSA is by undertaking PSG at a laboratory of sleep. It records the breath air flow, movement of respiratory, oxygen saturation, a position of the body, electromyography (EMG), electroencephalography (EEG) and electrocardiogram (ECG) for OSA detection. However, the high cost, long research time and inconvenience of patient testing of this technology make it difficult to widely used.

Over the past few years, various OSA detection methods were developed to extract information about frequency domain, time domain, and other nonlinear features from ECG signals to find the most representative features and then establish corresponding classifier. For example, found that many manuals extracted ECG features such as the heart rate (HR), S-pulse amplitude, and pulse energy are informative for OSA detection analyzed the Chazal and Yilmaz features of ECG signals via PCA transform proposed a novel method to detect OSA by extracted features from the QRS complex changes, which were caused by sleep apneas and spectral abnormalities of heart rate variability.

Various traditional neural network methods of OSA detection have been highly studied in recent years. proposed an OSA detection method based on MLP classifier. The algorithm used normalized Gray level co-occurrence matrix (NGLCM) to extract texture features, which were generated from images generated by short-time discrete Fourier transform (STDFT) of heart rate variability. And then seventeen extracted features were used as inputs to a three-level MLP classifier. proposed a novel OSA detection method by using neural network selected its best inputs for better identification of the trial data and identified the data during the feature matching step. The power spectral density of heart rate variability selected as input features were estimated by the soft-decision wavelet-based algorithm. applied an automated classification technique based on support vector machine to detect OSA. The method trained on a subject with and without OSA for training and testing the model which classification accuracy was up to 96.5%.

In the last decade, deep learning has experienced a great breakthrough and widely utilization. Traditional machine learning methods mainly use linear discriminant function to analyze and classify data. It is often difficult to achieve good classification results on a large number of complex and similar samples. Depending on its powerful autonomous learning ability and highly non-linear mapping, deep learning network can still achieve a good classification and identification results on some complex and high-precision classification problems. At present, it has been widely used in image classification, speech recognition, object detection, and so on.

II. LITERATURE SURVEY

Healthgear: Automatic Sleep Apnea Detection and Monitoring With A Mobile Phone

Nuria Oliver and Fernando Flores-Mangas - 2018

We present HealthGear, a real-time wearable system for monitoring, visualizing and analyzing physiological signals. HealthGear consists of a set of non-invasive physiological sensors wirelessly connected via Bluetooth to a cell phone which stores, transmits and analyzes the physiological data, and presents it to the user in an intelligible way. In this paper, we focus on an implementation of HealthGear using a blood oximeter to monitor the user's blood oxygen level and pulse while sleeping. We also describe two different algorithms for automatically detecting sleep apnea events, and illustrate the performance of the overall system in a sleep study with 20 volunteers.

Detection Of Obstructive Sleep Apnea From Cardiac Interbeat Interval Time Series

J.E. Mietus, C.K. Peng, P.Ch. Ivanov, A.L. Goldberger - 2017

Presents a new automated method to diagnose and quantify obstructive sleep apnea from single-lead electrocardiograms based on the detection of the periodic oscillations in cardiac interbeat intervals that are often associated with prolonged cycles of sleep apnea. This technique employs the Hilbert transformation of the sinus interbeat interval time series to derive the instantaneous amplitudes and frequencies of the series and calculates their averages and standard deviations over a moving 5-minute window. The authors then apply a thresholding technique and detect continuous sequences of those windows that lie within threshold limits. When applied to the Computers in Cardiology sleep apnea test data, the authors' algorithm correctly classified 28 out of 30 cases (93.3%) of both sleep apnea and normal subjects, and correctly identified the presence or absence of sleep apnea in 14,591 out of a total of 17,268 minutes (84.5%) of the data from the test set.

Online Obstructive Sleep Apnea Detection on Medical Wearable Sensors

Amir Aminifar, Francisco Rincón, Srinivasan Murali, David Aienza - 2019

Obstructive Sleep Apnea (OSA) is one of the main under-diagnosed sleep disorders. It is an aggravating factor for several serious cardiovascular diseases, including stroke. There is, however, a lack of medical devices for long-term ambulatory monitoring of OSA since current systems are rather bulky, expensive, intrusive, and cannot be used for long-term monitoring in ambulatory settings. In this paper, we propose a wearable, accurate, and energy efficient system for monitoring obstructive sleep apnea on a long-term basis. As an embedded system for Internet of Things, it reduces the gap between home health-care and professional supervision. Our approach is based on monitoring the patient using a single-channel electrocardiogram signal. We develop an efficient time-domain analysis to meet the stringent resources constraints of embedded systems to compute the sleep apnea score. Our system, for a publicly available database (PhysioNet Apnea-ECG), has a classification accuracy of up to 88.2% for our new online and patient-specific analysis, which takes the distinct profile of each patient into account. While accurate, our approach is also energy efficient and can achieve a battery lifetime of 46 days for continuous screening of OSA.

Automatic Detection And Quantification Of Sleep Apnea Using Heart Rate Variability

Saeedbaaeizadehphdadavid, P.Whitemdbcstephen, D.Pittmanmsbmedsophia, H.Zhouphd - 2019

Detection of sleep apnea using electrocardiographic (ECG) parameters is noninvasive and inexpensive. Our approach is based on the hypothesis that the patient's sleep-wake cycle during episodes of sleep apnea modulates heart rate (HR) oscillations. These HR oscillations appear as low-frequency fluctuations of instantaneous HR (IHR) and can be detected using HR variability analysis in the frequency domain. The purpose of this study was to evaluate the efficacy of our ECG-based algorithm for sleep apnea detection and quantification. The algorithm first detects normal QRS complexes and R-R intervals used to derive IHR and to estimate its spectral power in several frequency ranges. A quadratic classifier, trained on the learning set, uses 2 parameters to classify the 1-minute epoch in the middle of each 6-minute window as either apneic or normal. The windows are advanced by 1-minute steps, and the classification process is repeated. As a measure of quantification, the algorithm correctly classified 84.7% of all the 1-minute epochs in the evaluation database; and as a measure of the accuracy of apnea classification, the algorithm correctly classified all 30 test recordings in the evaluation database either as apneic or normal. Our sleep apnea detection algorithm based on analysis of a single-lead ECG provides accurate apnea detection and quantification. Because of its noninvasive and low-cost nature, this algorithm has the potential for numerous applications in sleep medicine.

Optimization Of Sleep Apnea Detection Using Spo2 And AN

Sheikh Shanawaz Mostafa, Joao Paulo Carvalho, Fernando Morgado-Dias, Antonio Ravelo-García - 2020

Repetitive respiratory disturbance during sleep is called Sleep Apnea Hypopnea Syndrome and causes various diseases. Different features and classifiers have been used by different researchers to detect sleep apnea. This study is undertaken to identify the better performing blood oxygen saturation features subset using an Artificial Neural Network classifier for sleep Apnea detection. A database of 8 subjects with one-minute annotation is used to test the proposed system. The optimized system has seven features chosen from a total set of sixty-one features presenting a high accuracy rate using a genetic algorithm. Artificial Neural Network was able to achieve 97.7 percentage of accuracy with only seven features chosen by the Genetic algorithm.

Problem Statement

Over the average lifespan, humans sleep for about 1/3 of their lives. A sleeping disorder is when one cannot sleep, causing the body to lose function. Just as the body's benefits of rest can range from physical to emotional and psychological effects, lack of sleep can damage the body physically, emotionally and psychologically. Till date, 84 kinds of sleep disorders have been discovered, including the most common ones such as insomnia, sleep apnea, narcolepsy and restless leg syndrome. Sleep Apnea (SA) is the instance when one either has pauses of breathing in their sleep, or has very low breath while asleep. This pause in breathing is known as an apnea, and can range in frequency and duration. The lack of breathing during sleep is known as a hypopnea. Sleep apnea is classified into two different types. The first type is Obstructive Sleep Apnea (OSA), which is more common, occurring in 2% to 4% of middle-aged adults and 1% to 3% of preschool children [3], and is generally caused by a collapse of the upper respiratory airway. The second one is Central Sleep Apnea (CSA), which is caused by an absent or inhibited respiratory drive. Most cases of CSA are mixed, meaning that it is often along with OSA cases, and the CSA only cases appear exceedingly rarely [4]. Despite how common it is, most cases go undiagnosed and can be attributed to 70 billion dollars loss, 11.1 billion in damages and 980 deaths each year [5]. Most sleep apnea cases go undiagnosed because of the inconvenience, expenses and unavailability of testing. The traditional testing process includes a polysomnography (PSG), which is a standard procedure for all sleep disorder diagnosis. It records the breath airflow, respiratory movement, oxygen saturation, body position, electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG).

III. METHODOLOGY

Proposed Work

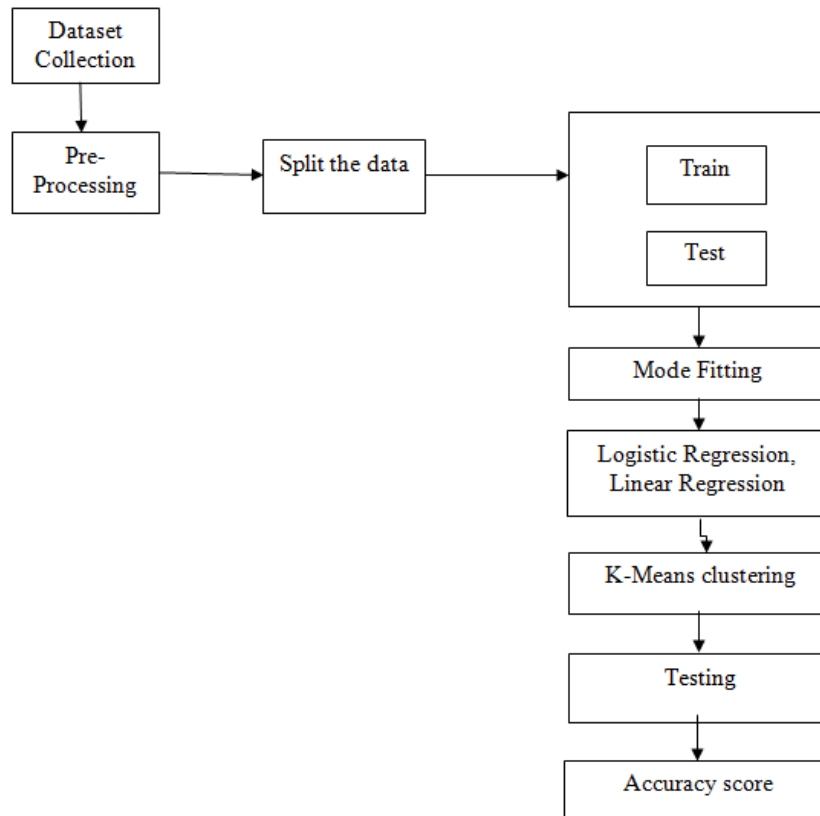
Our proposed system is to contemporary range of dataset classifiers used for the data set training of the machine learning algorithms (supervised and unsupervised learning) models include Logistic Regression, Linear Regression and K-Means clustering is proving to be effective in terms of detecting the accuracy of the systems. However, among the key areas wherein there is need for improvements in the accuracy improvements in the system could be attributed to more feature selections that could be considered.

The system if can take dynamic feature selection factors in to consideration based on the weighted average for specific conditions, the scope for detection of the OSA can be more emphatic. The scope of a new model could be about identification of the features in two stage model. The first stage could be about understanding the lifestyle and psychological conditions of the patient data and accordingly choose the metrics and the model of OSA detection tool that can be used for analysis.

If the proven metrics of the machine learning solutions are used, it can help even a routine diagnostic center with access to the application system can support in diagnosis conditions. One of the critical factors that is imperative in the condition is how the machine learning models are turning to be an effective solution in the case of the obstructive sleep apnea condition detection.

Advantages

- High Performance and accuracy.
- Easily Extract the Features.
- Easy detection of the Sleep Apnea disease.



3.1 Dataset Collection

A dataset (or data set) is a collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the dataset in question. It lists values for each of the variables, such as height and weight of an object. Each value is known as a datum.

We have chosen to use a publicly-available patient's data which contains a relatively small number of inputs and cases. The data is arranged in such a way that will allow those trained in disciplines to easily draw parallels between familiar statistical and novel ML techniques. Additionally, the compact dataset enables short computational times on almost all modern computers. Datasets are collected from Kaggle opensource website. That dataset includes patient detail and whose have sleep apnea disease or not.

3.2 Pre-Processing

In this step, we use various types of pre-processing techniques to handle the missing, noisy and inconsistent data. There are a number of pre-processing techniques such as case folding dam character erase, tokenization, slang word handling, stop word removal, stemming and number handling.

The sklearn.preprocessing package provides several common utility functions and transformer classes to change raw feature vectors into a representation that is more suitable for the downstream estimators.

3.3 Model Fitting

As mentioned before, our proposed system implemented on three algorithm that is logistic regression, Linear regression and K-Means clustering. In K-means the number of clusters is already specified prior to running the model. We can choose a base level number for K and iterate to find the most optimum value. To evaluate which number of clusters is more optimum for our dataset, or find *cluster fitness* we use two scoring methods — **Silhouette Coefficient** and **Calinski Harabasz Score**. In reality, there are many different scoring methods depending on what metrics matter most in a model. Usually, one method is chosen as the standard but for the purpose of this analysis I have used two.

The Silhouette Coefficient is calculated using the **mean intra-cluster distance (a)** and the **mean nearest-cluster distance (b)** for each sample. The Silhouette Coefficient for a sample is $(b-a) / \max(b-a)$

The Calinski Harabasz Score or Variance Ratio is the ratio between **within-cluster dispersion** and **between-cluster dispersion**. Let us implement the K-means algorithm using sci-kit learn. n_clusters= 12

3.4 Prediction

We can see that for both types of scores, 8 clusters give a better value. However, we will have to do several iterations with different number of clusters to find the optimal one. Instead, we can use something called an **elbow plot** to find this optimal value. An elbow plot shows at what value of k, the distance between the mean of a cluster and the other data points in the cluster is at its lowest. Two values are of importance here — *distortion and inertia*. Distortion is the average of the euclidean squared distance from the centroid of the respective clusters. Inertia is the sum of squared distances of samples to their closest cluster centre. Here we can see the drop in the sum of squared distance starts to slow down after k=5. Hence 5 is the optimal number of clusters for our analysis. We can verify this by calculating the silhouette coefficient and CH score for k=5.

IV. IMPLEMENTATION

TABLE 1
 DATASET

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

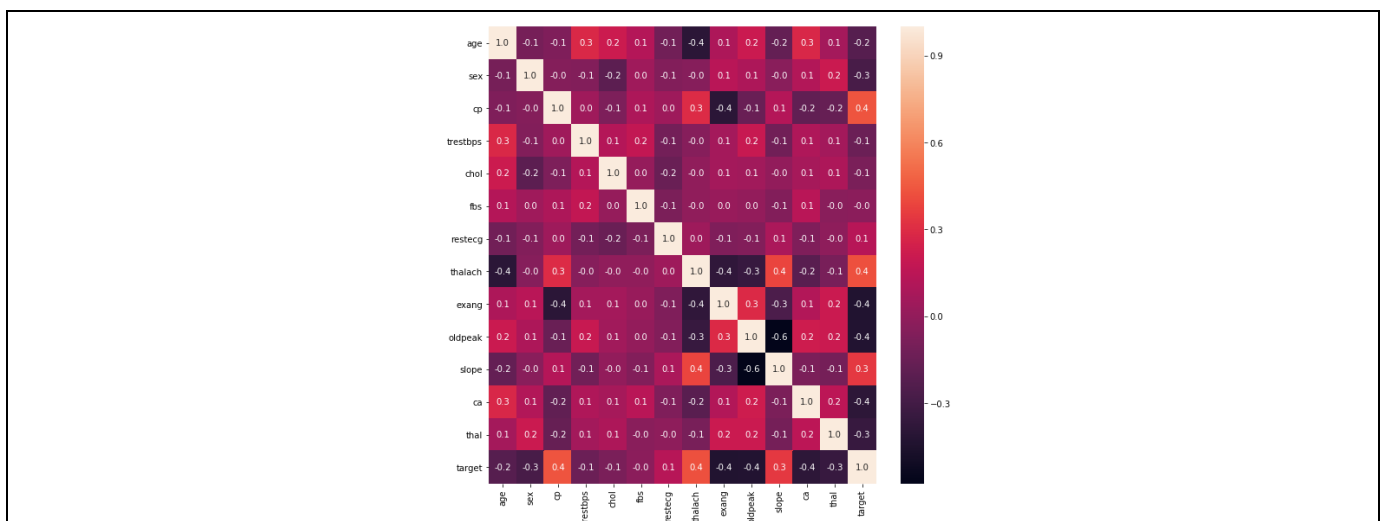


FIGURE 1: Heat Map

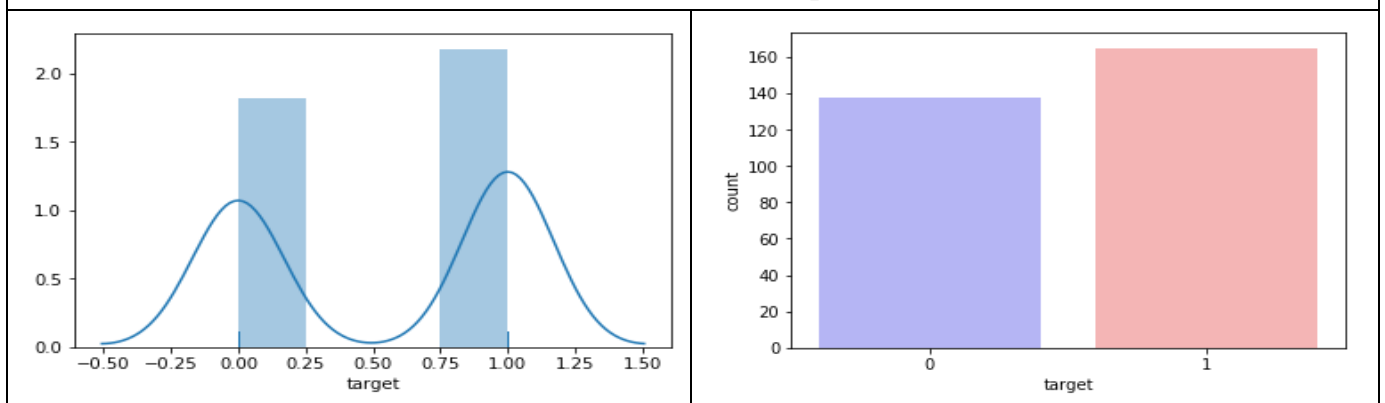


FIGURE 2: Data Visualization : Target and Count

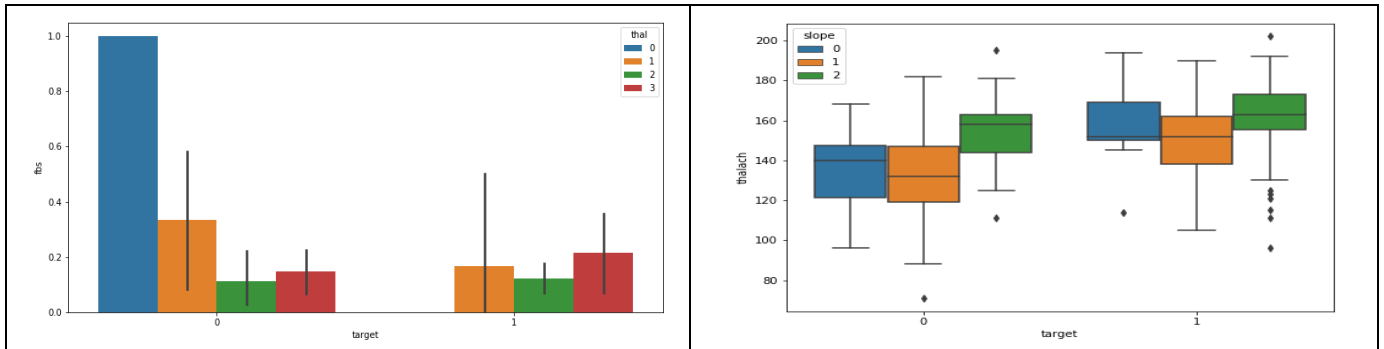


FIGURE 3: Categorical_variable: target, fbs and thalach

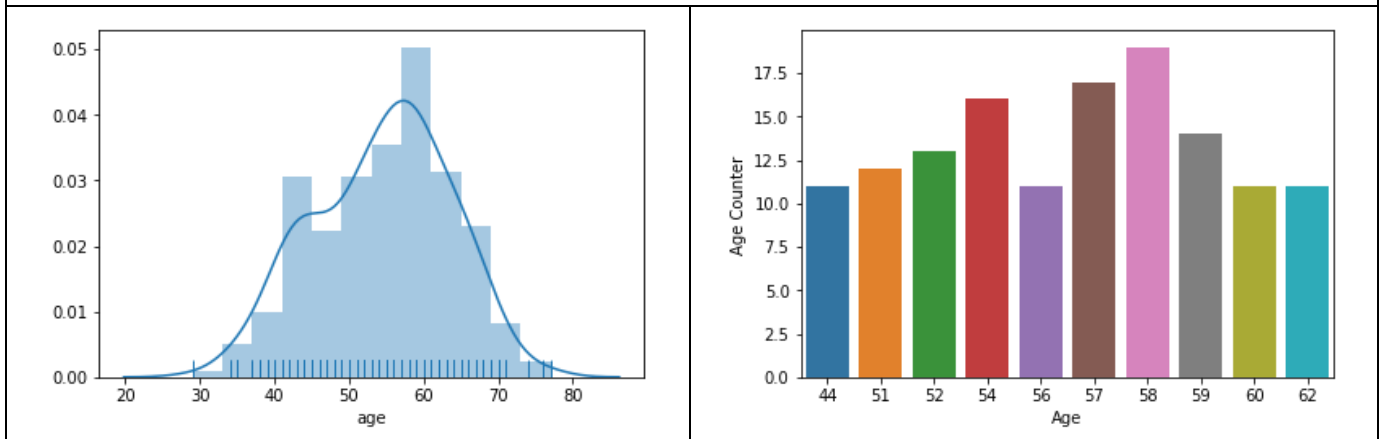


FIGURE 4: Categorical_variable – age and age counter

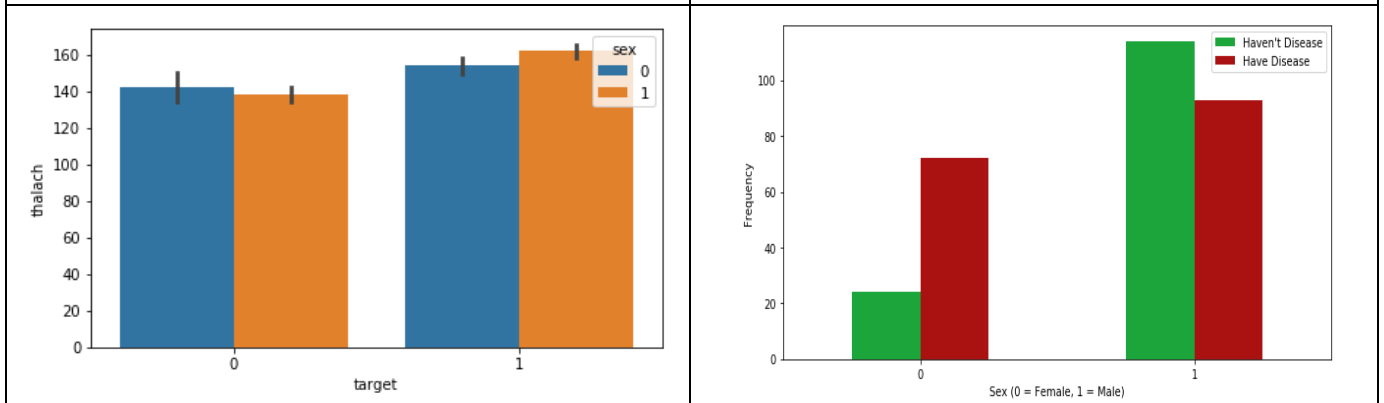
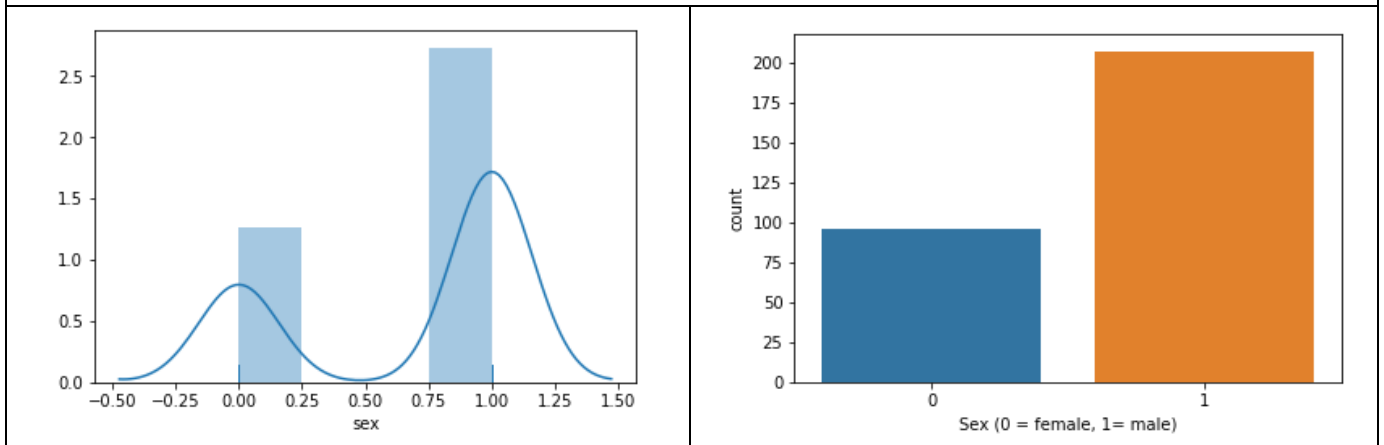


FIGURE 5: Women are 4 times more likely to die from heart disease than breast cancer

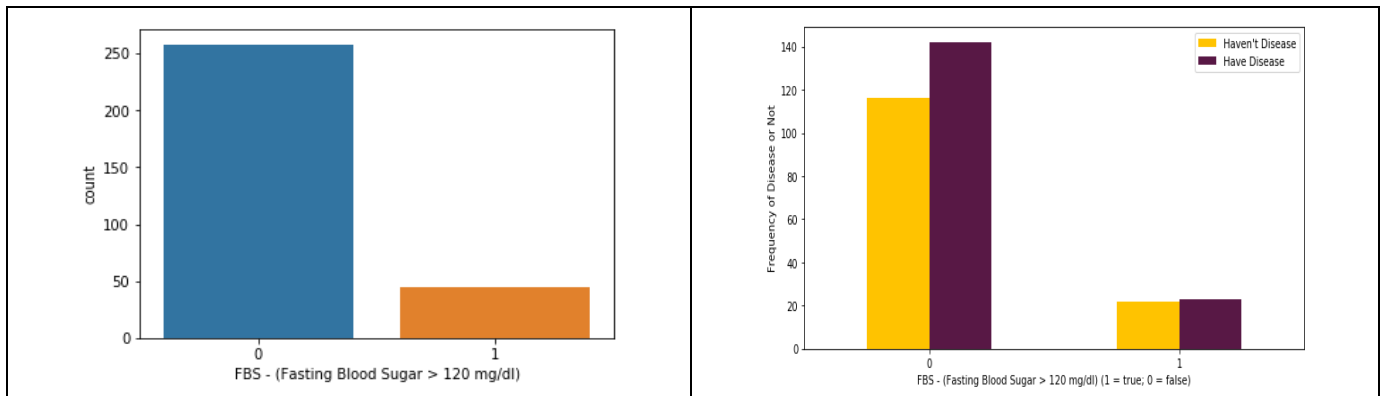


FIGURE 6: Fasting blood sugar

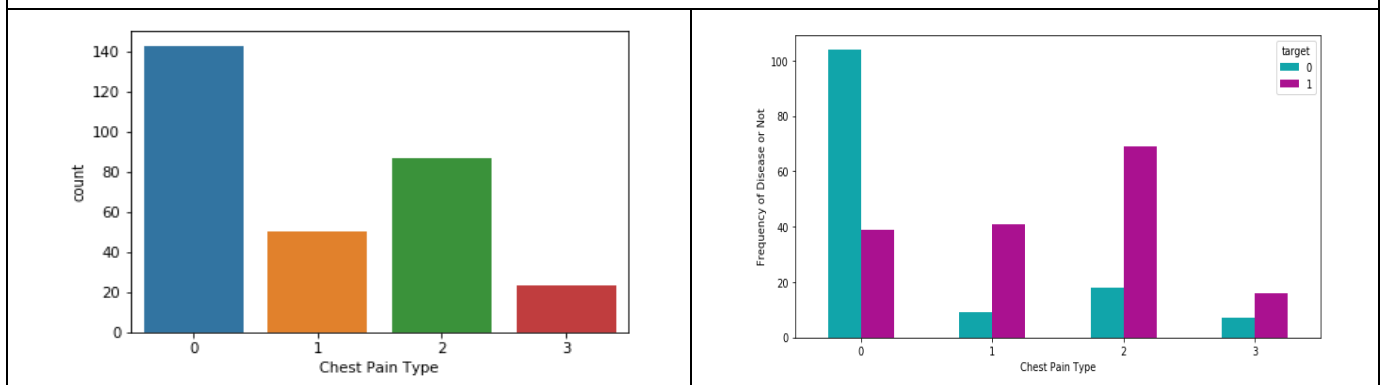


FIGURE 7: Chest pain, count and frequency of disease or not

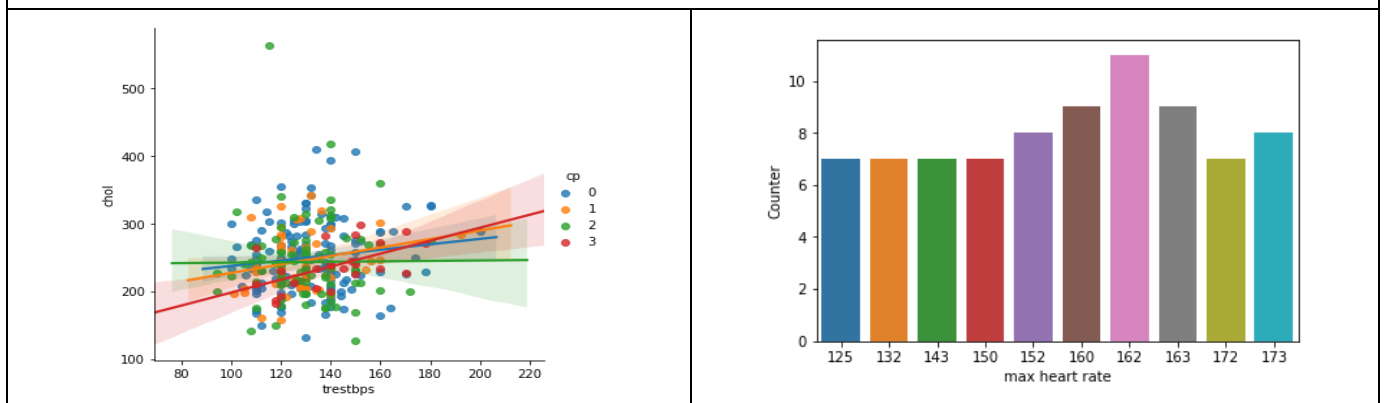


FIGURE 8: Thalach: maximum heart rate achieved

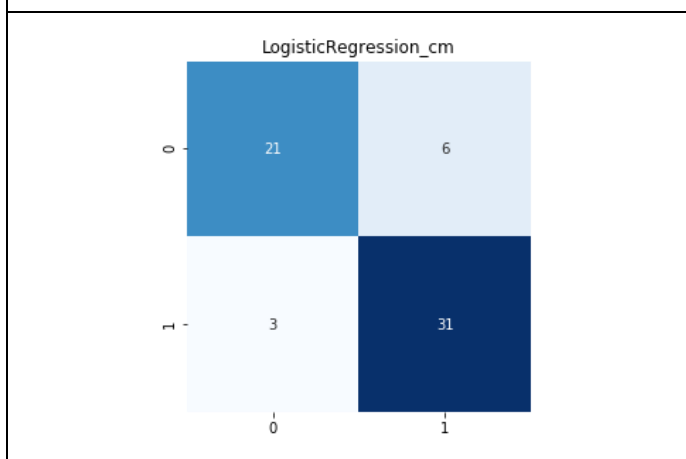


FIGURE 9: Confusion matrix for accuracy

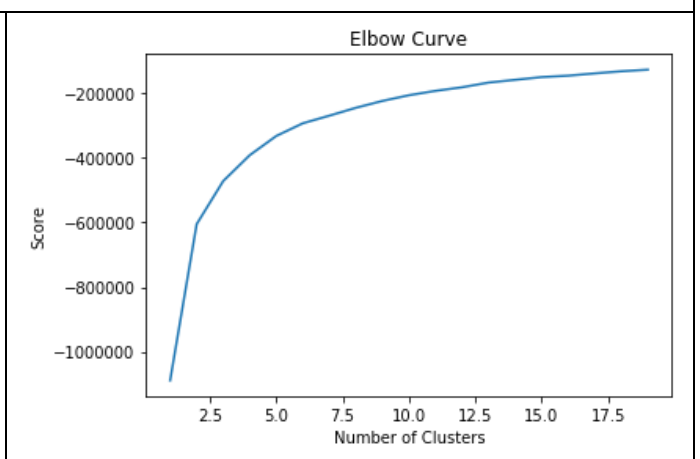
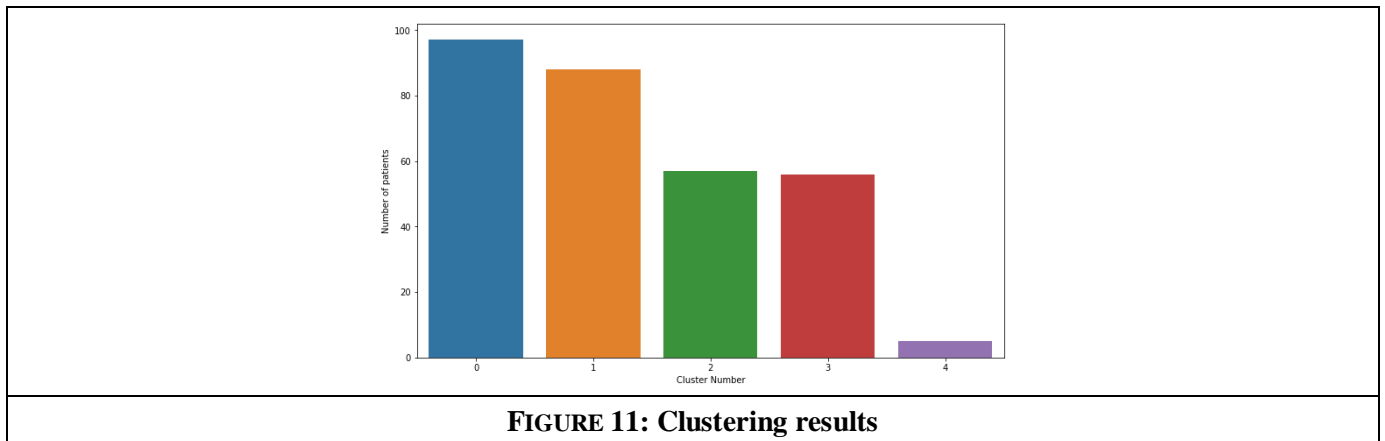


FIGURE 10: Fit K-Means clustering for k=5



V. CONCLUSION

OSA is becoming one of the significant challenges encountered by the public and there are increasing percentage of cases reported for the problem. Profoundly classified as central apnea, obstructive sleep apnea and mixed sleep apnea conditions. Till date, one of the keyways in which the problem is diagnosed in a methodical system is about testing the patient for polysomnography tests, which takes in to account distinct conditions and to identify the AHI conditions. But considering the quantum of resource requirement and the time consumption in the process, there are many alternative studies and models that are evident in the literature. Towards enhancing the quality of such diagnostic solutions, the machine learning based solution models are too prevalent. This manuscript has reviewed distinct set of alternate models and the machine learning models that were proposed, which can help in improving the accuracy of the diagnostics. However, majority of the models that were proposed in the machine learning approach is selection of health functional parameters as metrics and features. Whereas, there is possibility of choosing the symptoms too as the metrics, which can be easier to focus for detection, as the data can be more easily garnered from the patients. If a conglomerated model of multi-tiered approach for detection if designed too, can be an effective model for development.

REFERENCES

- [1] Xie, B., & Minn, H. (2012). Real-time sleep apnea detection by classifier combination. *IEEE Transactions on information technology in biomedicine*, 16(3), 469-477.
- [2] Nandakumar, Rajalakshmi, Shyamnath Gollakota, and Nathaniel Watson. "Contactless sleep apnea detection on smartphones." In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pp. 45-57. 2015.
- [3] Mendonca, F., Mostafa, S.S., Ravelo-Garcia, A.G., Morgado-Dias, F. and Penzel, T., 2018. A review of obstructive sleep apnea detection approaches. *IEEE journal of biomedical and health informatics*, 23(2), pp.825-837.
- [4] Oliver, Nuria, and Fernando Flores-Mangas. "Healthgear: Automatic sleep apnea detection and monitoring with a mobile phone." *JCM* 2, no. 2 (20017): 1-9.
- [5] Yadollahi, A. and Moussavi, Z., 2009, September. Acoustic obstructive sleep apnea detection. In *2009 annual international conference of the IEEE engineering in medicine and biology society* (pp. 7110-7113). IEEE.
- [6] Mietus, J.E., Peng, C.K., Ivanov, P.C. and Goldberger, A.L., 2000, September. Detection of obstructive sleep apnea from cardiac interbeat interval time series. In *Computers in Cardiology 2000*. Vol. 27 (Cat. 00CH37163) (pp. 753-756). IEEE.
- [7] Hassan, A.R., 2016. Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting. *Biomedical Signal Processing and Control*, 29, pp.22-30.
- [8] Song, C., Liu, K., Zhang, X., Chen, L. and Xian, X., 2015. An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals. *IEEE Transactions on Biomedical Engineering*, 63(7), pp.1532-1542.
- [9] Nguyen, H.D., Wilkins, B.A., Cheng, Q. and Benjamin, B.A., 2013. An online sleep apnea detection method based on recurrence quantification analysis. *IEEE journal of biomedical and health informatics*, 18(4), pp.1285-1293.
- [10] Almazaydeh, L., Elleithy, K. and Faezipour, M., 2012, August. Obstructive sleep apnea detection using SVM-based classification of ECG signal features. In *2012 annual international conference of the IEEE engineering in medicine and biology society* (pp. 4938-4941). IEEE.