Research Paper: Computational Aspects and Statistical Models in Sleeping Disorders; an Apriori Algorithm Approach

Mani Butwall[1], Kinshuk Gaurav Singh[1], Raj Pujara[1], Pranav Modi[1], Harshvardhan Sharma[1], Arsh Vishwakarma[1], Iman Salehi[2], Mohammad Javad Gholamzadeh[3], Ali-Mohammad Kamali[2,4], Milad Kazemiha[2,4], Prasun Chakrabarti[5], Mohammad Nami[2,4,6*]

1. Department of Computer Science, ITM (SLS) Baroda University, Vadodara, India
2. DANA Brain Health Institute, Iranian Neuroscience Society-Fars Chapter, Shiraz, Iran
3. Students' Research Committee, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
4. Department of Neuroscience, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran
5. Provost, Techno India JNR, Institute of Technology, Udaipur 313003, Rajasthan, India
6. Neuroscience Center, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT AIP), City of Knowledge, Panama City, Republic of Panama

Use your device to scan and read the article onlinne

Citation Butwall M, Gaurav Singh K, Pujara R, Modi P, Sharma H, Vishwakarma A, Salehi I, Gholamzadeh MJ, Kamali AM, Kazemiha M, Chakrabarti P, Nami M. Computational aspects and statistical models in sleeping disorders; an apriori algorithm approach. JAMSAT. 2021; 6(1):5-13.

https://dx.doi.org/10.30476/jamsat.2021.93286.1025

Article info:
Received: 17 October 2021
Accepted: 17 November 2021

Keywords:
Sleep Disorders, Statistical Analysis, Apriori Algorithm, Simple K-Means Clustering, Random Forest Regressor, Pearson correlation coefficient, Odds ratio, Principal Component Analysis

ABSTRACT

Sleep disorders are very common in today's world as we all are living a relatively competitive life, where we undervalue our mental health. There are some sleep disorders that share almost similar symptoms yet various pathological underpinnings leading to clinical misjudgments, thereby resulting in the inappropriate treatments. The present study has attempted to investigate possible correlation between various types of sleep predicaments. To do so, we used multiple statistical analysis algorithms as well as prediction models on our database and performed manual testing to draw our conclusion. Our analyses revealed that sleep disorders, and namely sleep apnea-hypopnea syndrome, tend to present with related comorbidities.

*** Corresponding Author:**
Mohammad Nami, PhD
Address: *Department of Neuroscience, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran*
Tel: *+987132305488*
E-mail: *torabinami@sums.ac.ir*

1. Background

Sleep disorders are a group of conditions that affect the ability to sleep well on a regular basis, whether they are caused by a health problem or by too much stress. Depending on the type of sleep disorder, people may have a difficult time falling asleep and may feel extremely tired throughout the day. The lack of sleep can have a negative impact on energy, mood, concentration, and overall health. But sometimes it is possible that having symptoms of one sleeping disorder can lead to another sleep disorders. There are several different types of sleep-wake disorders, of which insomnia is the most common. Other sleep-wake disorders include obstructive sleep apnea, parasomnias, narcolepsy, and restless leg syndrome.

A. Apriori Algorithm

Apriori is an algorithm for frequent item set mining and association rule learning over relational databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database.

B. Simple K means Clustering

K-Means Clustering is a method of vector quantization, originally from signal processing that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster.

C. Random Forest Regressor

A random forest is a Meta estimator that fits a number of classifying decision trees on various sub-sample of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

D. Pearson Correlation Coefficient

It is used to measure the linear correlation between two random variables to have a clear picture on variability and interpretability. The formula is:

$$r = \frac{1}{n-1}\sum_{i=1}^{n} \frac{(-m_x)(y_i - m_i)}{s_x * s_y}$$

Where,

r = Pearson coefficient, x = variable 1, y = variable 2, m_x = Mean of variable x, m_y = Mean of variable y, N = No. of samples, S_x = Standard Deviation of x, S_y = Standard Deviation of y

It gives a value between -1 & 1 such that;

If r = 0, no correlation

$0 < r < 1$, positive correlation

$-1 < r < 0$, negative correlation

E. Odds Ratio

It defines whether there is an association present between two properties or not. It can act as a proxy but not as a statistical inference.

If OR (Odds Ratio) is exactly 1, that means first property does not affect the second. Higher than 1 denotes higher odds of second property with exposure to first property.

As we know that OR is not statistically significant. To find the precision of odds ratio, we used 95% CI (confidence interval).

F. Principal Component Analysis

It is a technique used for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. It does so by creating new uncorrelated variables that successively maximize variance.

2. Related Work on the Modelling of Sleep Disorders

A. Sleep Disorder

Changes in sleeping patterns or habits that can negatively affect health can be addressed as Sleep Disorders.

B. Types of Sleep Disorders

Insomnia: It is one of the common but neglected conditions seen in family practice with long term and serious effects on the health of a patient (1). Cognitive Behavioral Therapy for Insomnia (CBT-I) is the most prominent non-pharmacologic treatment for insomnia disorders (2). Although meta-analyses have examined primary insomnia, less is known about the comparative efficacy of CBT-I on comorbid insomnia (3).

Parasomnia: Although sleepwalking (somnambulism) affects up to 4% of adults, its pathophysiology remains poorly understood (4).

Nami M, et al. The role of bihemispheric cerebellar anodal transcranial electrical stimulation in functional outcome among strength athletes. JAMSAT. 2021; 6(1):5-13..

Sleepwalking can be preceded by fluctuations in slow-wave sleep Electroencephalography (EEG) signals, but the significance of these pre-episode changes remains unknown and methods based on EEG functional connectivity have yet to be used to better comprehend the disorder (5).

Table 1. A survey on sleep disorders

Researcher	Dataset	Samples	Methodology	Accuracy (%) / Result
Bhaskar et. al. (2016) (1)	Collected data was alphanumerically coded and entered in an Excel sheet.	278 adult patients attending the OPD from September 1 to October 30, 2015	The analysis was done using SPSS 19.0 software (IBM, Bangalore, India).	Insomnia is a common sleep disorder which is many times missed by a primary care physician until/unless asked for; the prevalence of which was as high as 33% in this study.
Wu et. al. (2015) (2)	A systematic search terms included (1) CBT-I or CBT or cognitive behavioral (and its variations) (2) insomnia or sleep disturbance.	37 Studies were included.	CBT-I studies were included.	36.0% of patients who received CBT-I were in remission from insomnia compared with 16.9% of those in control or comparison conditions (pooled odds ratio, 3.28; 95% CI, 2.30-4.68; P < .001).
park et al. (2019)(3)	Sleep pattern (Fitbit), Daily activity (Fitbit), Personal demographic (Survey)	50 samples out of which 8 were excluded because they didn't wear the device for longer than 2 days.	Diachronic unsupervised learning of insomnia patterns and then clustering individuals with insomnia	This finding means that our neural net–based clustering method could identify, beyond conventional diagnosis, new meaningful sleep-activity relationships that could be used to devise tailored interventions. Method could determine which cluster an individual belongs to via data indicators of sleep and behavior acquired in the study.
Laberge et al. (2000) (4)	Data collected during the course of a longitudinal study of a representative sample of children from Québec.	The present analyses are based on results available for 664 boys and 689 girls for whom mothers have completed questions concerning demographics, parasomnias, and anxiety level.	For the prevalence and developmental aspects of parasomnias, prospective data were collected at annual intervals from 11 to 13 years old and retrospective data for the period between ages 3 and 10 years were collected when the children were 10 years old.	Although sleepwalking, night terrors, enuresis, and body rocking dramatically decreased during childhood, somniloquy, leg restlessness, and sleep bruxism were still highly prevalent at age 13 years, paralleling results found in adults.
Desjardins et al. (2020) (5)	EEG of 27 adult sleepwalkers; The 20-second segment of sleep EEG immediately preceding each patient's episode was compared with the 20-second segment occurring 2 minutes prior to episode onset.	27 adult sleepwalkers (13 men, 14 women, mean age: 29 ± 7.6 years).	Statistical analysis (Paired t-Test)	P > 0.05

JAMSAT
Journal of Advanced Medical Sciences and Applied Technologies

Table 1. Continue

Researcher	Dataset	Samples	Methodology	Accuracy (%) / Result
Bollu et al. (2019) (6)	-	-	This article briefly discusses the clinical characteristics, demographics, and pathophysiology of major parasomnias and associated disorders.	RBD has shown to be strongly associated with an underlying neurodegenerative α-synucleinopathy while many other parasomnias are benign.
Osman et al. (2016) (7)	EEG dataset	-	comprehensive in-laboratory polysomnography	New simplified approaches to estimate each of the key causes of OSA been developed
Vahabi et al. (2021) (8)	EIT datasets	-	15 premature neonates with an episode of apnea in their breathing pattern	hybrid classification model consisting of convolutional neural network with 50 layers deep (ResNet50) architecture, and a Support Vector Machine (SVM) classifier
Kendzerska et al. (2014) (10)	clinical database, health administrative data	-	Cox regression models and polysomnography	OSA-related factors other than AHI were shown as important predictors of composite CV outcome.
Stephansen et al. (2018) (11)	-	-	The Stanford-stages app uses ML to perform automated sleep stage scoring and narcolepsy to identify nocturnal PSG studies.	A T1N marker based on unusual sleep stage overlaps achieved a specificity of 96% and a sensitivity of 91%, validated in independent datasets.
Zhang et al. (2018) (12)	EU-NN database.	1380 patients where, 702 narcolepsy patients (598-NT1, 104-NT2, 183-hypocretin deficiency) & 678 Healthy controls	Stochastic Gradient Boosting (SGB) model	Accuracy of prediction 0.9943 (95% CI: 0.9686 - 0.9999)
Exarchos et al. (2019) (13)	Time-locked EEG, EMG, and infrared video of freely behaving mice from The Jackson Laboratory were simultaneously recorded.	*n* = 7, 6 females, adult aged 3–6 months, and wild-type mice where 3 hemizygous transgenic narcoleptic mice.	Supervised learning-CNN, SVM Unsupervised learning (tSNE + Density-Based Spatial Clustering of Applications with Noise (DBSCAN))	Mean accuracy of 95% and 91% in narcoleptic mice where CNN exhibits superior performance compared to the SVM/unsupervised methods.
Yıldız et al. (2018) (14)	-	70 samples, in which 35 are with RLS and other 35 are healthy.	The correlation between the severity of RLS symptoms and HRV parameters measured on an electrocardiogram was analyzed.	A decreased SDNN, SDANN, and SDNN index, and an increased LF power and LF/HF ratio may be the early signs of cardiac autonomic dysfunction in patients with RLS.
Rahimde et al. (2012) (15)	-	65 patients with primary RLS	The 3 periods of drug prescription with one month of wash out period as placebo, 50 µg and 200 µg of selenium were administered in each separated month.	Improvements is seen in patients with selenium (50 and 200 µg) than placebo group.
Iftikhar et al. (2020) (16)	-	300 cases of diabetes mellitus type-2. Data was analyzed by using IBM SPSS version 23.	Point-biserial correlation/Pearson Chi-Square correlations were conducted between RLS and risk factors.	RLS was found in a significant association with smoking (p=0.045), hypertension (p=0.047) and chronic renal failure (CRF) (p=0.027).

Nami M, et al. The role of bihemispheric cerebellar anodal transcranial electrical stimulation in functional outcome among strength athletes. JAMSAT. 2021; 6(1):5-13..

Sleep Apnea: Obstructive Sleep Apnea (OSA) is an increasingly common, chronic, sleep-related breathing disorder. OSA is characterized by periodic narrowing and obstruction of the pharyngeal airway during sleep (6). Apnea in adults is a breathing disorder event that occurs as a result of the absence of inspiratory airflow for at least 10 seconds (7).

Narcolepsy: It is a rare life-long disease that exists in two forms, Narcolepsy Type-1 (NT1) or Type-2 (NT2), but only NT1 is accepted as clearly defined entity (8). Both types of narcolepsies belong to the group of Central Hypersomnia (CH), a spectrum of poorly defined diseases with excessive daytime sleepiness as a core feature (9).

Restless Leg Syndrome (RLS): It is defined as an uncomfortable feeling in the limbs which is prominently sensed in legs (10). Dopamine system involvement is considered as the basis of RLS's etiology (11).

3. Literature Review on Different Types of Sleep Disorders

In this section, the survey of sleep disorders using various approaches is presented. Table 1 summarizes the various approaches used in literature.

Insomnia is a common sleep disorder which is many times missed by a primary care physician until/unless asked for; the prevalence of which was as high as 33% (1). A small to medium positive effect was found across comorbid outcomes, with larger effects on psychiatric conditions compared with medical conditions (2). Park et al. suggests that unsupervised learning allows health practitioners to devise precise and tailored interventions at the level of data-guided user clusters (i.e., precision psychiatry), which could be a novel solution to treating insomnia and other mental disorders (3).

Parasomnias are highly prevalent in children between the ages of 3 and 13 years (4). Desjardins et al. study confirms the presence of high anxiety levels in children suffering from night terrors and body rocking and reports it for the first time in children afflicted with somniloquy, leg restlessness, and bruxism in a large, controlled epidemiologic study. Results suggest that somnambulistic episodes are preceded by changes in brain processes that are relatively gradual in nature and that the interplay between sleep and wakefulness can be observed through EEG functional connectivity networks before the onset of such clinical events (5). It is important for the sleep physician to understand the nature of the parasomnia and to differentiate it from nocturnal seizures. Understanding the phenomenology of parasomnias is important due to the forensic implications associated with them (6).

Characterization of the different causes or phenotypes of OSA in recent years has provided new pathways for targeted therapy (7). Hybrid transfer learning technique has been proposed to classify apnea and non-apnea events with an accuracy of between 71% and 97% when the testing dataset is applied (8). OSA-related factors other than AHI were shown as important predictors of composite CV outcome and should be considered in future studies and clinical practice (9).

Zhang et al. show that machine learning approach may be valuable in diagnosing subtypes of narcolepsy and selecting new clinical features of narcolepsy on machine scale (11). The CNN classifier exhibits superior performance compared to the SVM and the unsupervised methods (12).

A decreased SDNN, SDANN, and SDNN index, and an increased LF power and LF/HF ratio may be the early signs of cardiac autonomic dysfunction in patients with RLS (12). Selenium prescription in daily recommended dose of 50 µg instead of a dopamine against would be an alternative treatment in improvement of RLS symptoms (13). There is a need to ascertain the incidence of RLS in diabetes patients to save them from comorbidities through more advanced tools of artificial based models (14).

4. Description of Database

Database contains details of 100 patients suffering from sleep disorders e.g. Hypopnea and Apnea.

Each Patient has around 800 -1000 epochs for more accuracy and result for our study.

Following details were Noted Down for each Patient:

Age, Sex, Weight, Height, Body Mass Index (BMI), Weight Status, Diagnosis , Sleeping States and Lastly Disorder Events. note that by "Events", we mean that multiple events for multiple sleeping disorders were observed to conclude our Study.

Events contains the following Abbreviation: AR = Arousal, CA = Central Apnea, Awake = Awake State, OH = Obstructive Hypopnea, MH = Mixed Hypopnea, OA = Obstructive Apnea, MA = Mixed

Apnea, REM Aw = Rapid Eye Movement (Awake), MChg = Montage Change, PLM = Periodic Leg/ Limb Moment, CH= Central Hypopnea, LM = Leg Movement.

5. Method

While working with the dataset, we had data of 100 patients with sleep disorders. After studying, 62 out of 100 patients were identified to be suffering from sleep apnea syndrome. Hence we focused on finding the correlation between sleep apnea syndromes with respect to other sleep disorders/diseases. Two inferences were made;

(1) At first, we categorized the patients according to symptomatic features, where CA, MA and OA implies that the patient is suffering from sleep apnea & CH, MH and OH implies that the patent is suffering from sleep hypopnea. According to the few literature survey that we did, we found that patients with apnea of 10 seconds or more, and more than 5 times per hour are confirmed for a diagnosis with sleep disorders (16). If someone experiences five or more hypopnea events per hour of sleep, they are likely to have a hypopnea sleep disorder (17).

We segregated the event files of each and every patient having recordings of their sleep cycles into 120 epoch slots (i.e., 1 hour), to check whether the patient was suffering from particular sleep disorder. Primarily, sleep apnea, sleep hypopnea or both.

After observing the frequency, we tried to find correlation between sleep apnea and sleep hypopnea. For that we first noted down occurrence of events in binary representation.

After labelling down all of the patients with either 0 or 1 we find odds ratio with 95% CI (confidence interval) of the test. As odds ratio is not statistically significant result, we considered Pearson correlation as our conclusive result.

(2) In the second inference, Apriori Algorithm was used, as join and Prune steps of algorithm can be easily implemented on larger dataset. The data was pre-processed and dimensionality reduction was performed. Numerical attributes, presented in our dataset, were discarded and only the nominal attributes (like weight Status, Diagnosis, and Medication etc.) were considered. Applying Apriori algorithm on dataset provided no conclusive evidence within acceptable confidence and support. So we used filter association rule was used as Apriori as base Associator and made some changes to Apriori algorithm. We achieved significant results with confidence level up to 0.9

Through this process Apriori algorithm was able to generated 5 item sets and on the basis of those item sets 10 rules were generated. Algorithm was able to find association between Apnea, overweight and depression with confidence of 0.94.

Then we used Simple K-means clustering to find the same relation that we found using Apriori algorithm. In Figure 5.1.1 plotting it is clearly visible that OSAS was heavily correlating with depression.

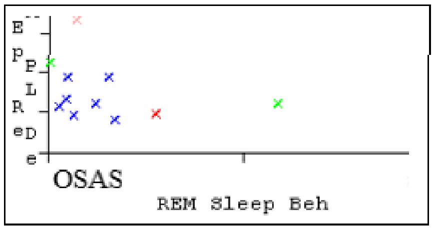

Figure 5.1.1. Relation between OSAS and depression

In Figure 5.1.2, we were able to observe that overweight/obese patients diagnosed with OSAS had high probability of being depressed.

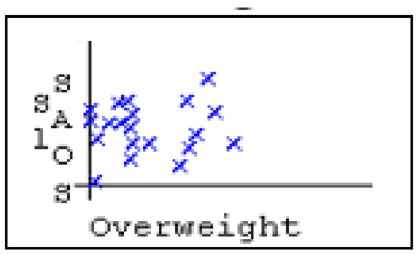

Figure 5.1.2. Relation between OSAS and depression

(3) After the first 2 tests, we then created a prediction model which was able to predict patients' diagnosis as well as other diseases on the basis of height, weight with the accuracy of 95.85% and 98.86% respectively.

For that we first loaded the dataset in python then performed following steps accordingly:

a) Pre-Processing of data

b) Checking data for normalization and standardization but similar results were observed. Therefore, applied Min-Max Normalization. After the above steps we applied PCA (Principal Component Analysis) to check which element was helping in prediction.

c) In the next step, we applied Random Forest Regressor for the training of 100 subjects and to test on 26 subjects to predict diagnosis and other problems.

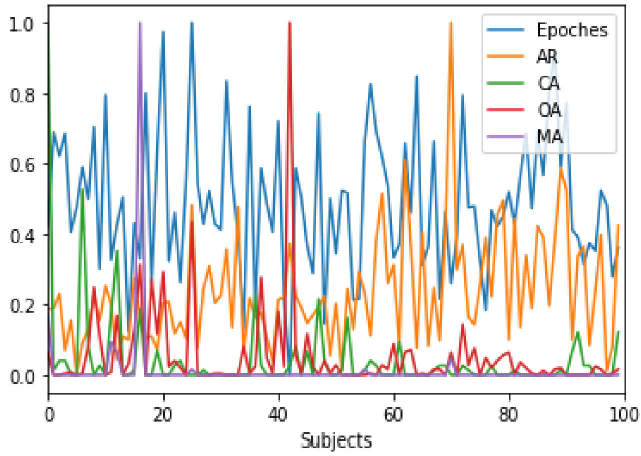

Figure 5.1.3. Relation between AR, OA, CA, and Epochs

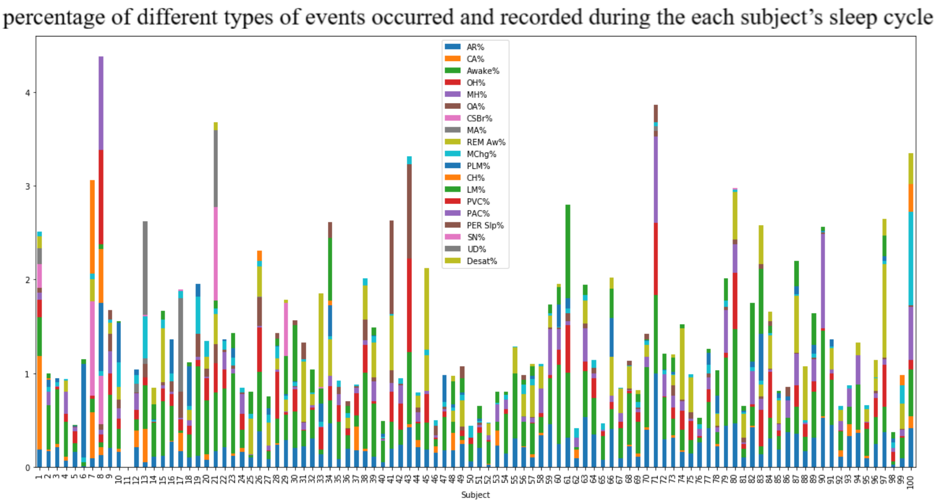

Figure 5.1.4. Shows the percentage of different types of events occurred and recorded during the each subject's sleep cycle

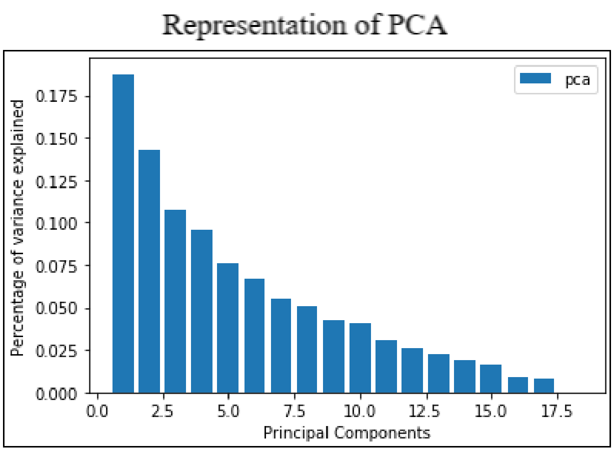

Figure 5.1.5. Represents Principal Component Analysis (PCA) graph that shows which element is helping in predictions

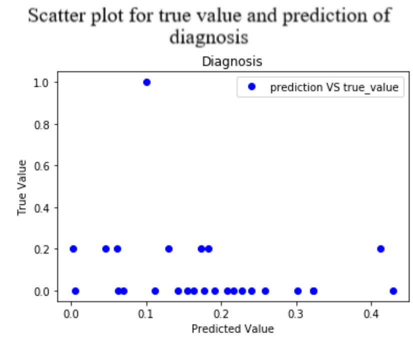

Figure 5.1.6. Shows Scatter plot for true value and prediction of diagnosis

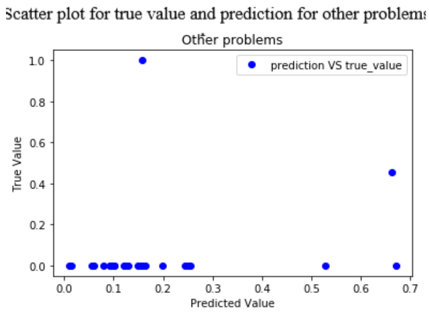

Figure 5.1.7. Represents Scatter plot for true value and prediction for other problems

6. Results

For the statistical analysis done on weka, the result was that out of all OSAS patients, 70.9677% were over-weight or obese and 8.4545% were suffering from depression.
The odds ratio shown that odds of carrying the sleep hypopnea variant were higher in those with sleep apnea, compared with those without sleep apnea (OR = 12.83; 95% CI = 4.45, 37.04).

	Hypopnea (Yes)	Hypopnea (No)
Apnea (Yes)	56	6
Apnea (No)	16	22

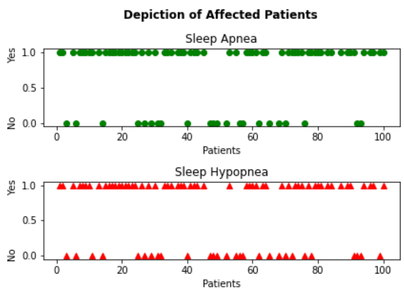

Figure 5.2.1. Shows the depiction of patients affected with sleep apnea and sleep hypopnea using binary matrix where 1 means Yes & 0 means No. Pearson correlation was 0.85 which strongly signifies that the occurrence of sleep hypopnea is positively correlated to sleep apnea

As for the Prediction model, we were able to successfully predict patient's diagnosis and other symptoms with the accuracy of 95.85% and 98.86%, respectively.

	diagnosis	Other symptoms
Mae	0.205	0.194
Mape	4.155	1.142
Rmse	0.267	0.273
SD of prediction	0.111	0.17
CoV of prediction	0.614	0.916

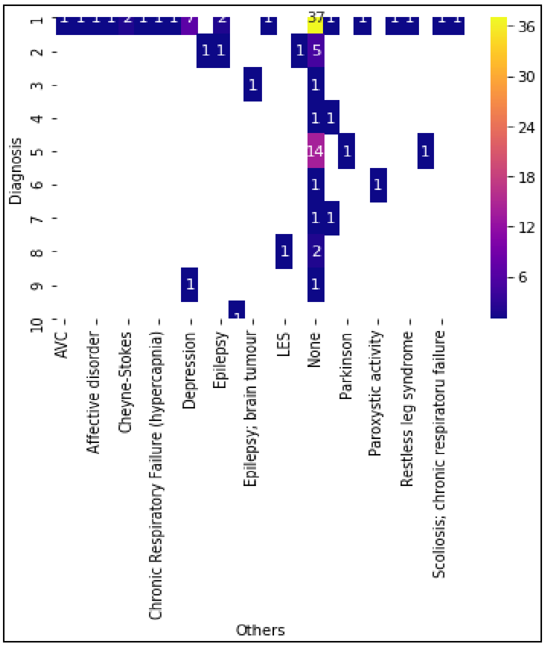

Figure 5.2.2. Heat-map showing relation between diagnosis and other problems

From the above figure 5.2.2, we can see that many patients with apnea (~63%) have some other kinds of problem majorly depression and some respiratory disease but others (~37%) don't have any other problem. Also, patients with snoring usually do not have any secondary problem except minor (~12.5%) complaints.

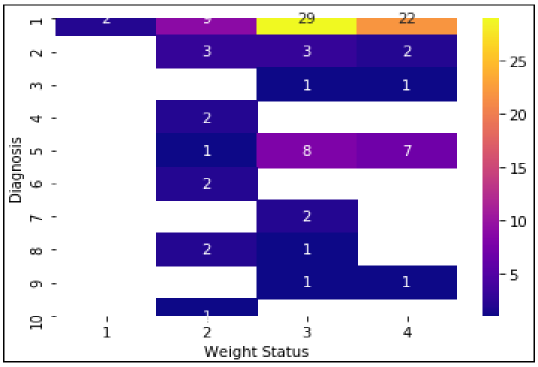

Figure 5.2.3. Heat-map of weight status and diagnosis. The annotations for diagnosis include 1: OSAS, 2: D. Affective, 3: PLMS, 4: REM sleep behavior disorder, 5: Snoring, 6: Epilepsy, 7: RLS, 8: Sleep deprivation, 9: Parasomnia, 10: SRVAS. The weight status is categorized as: 1=underweight, 2=healthy weight, 3=overweight and 4=obese. This figure shows that the mostly the apnea patients are overweight (49%) or obese (35%), together 89% and patients with snoring are also 50% are overweight or 43.75% obese, together 93%

7. Conclusion

We were able to come to conclusion that there was an association between sleep apnea and hypopnea events. Also, depression was found in apnea patients who were obese/overweight.

As for the prediction model, we were able to predict diagnosis as well as other diseases for each patient and concluded that apnea patients were also suffering from some other comorbid symptoms related to other types of sleep disorders. Through all the tests, we conclude that sleep disorders, and namely sleep apnea-hypopnea syndrome tend to present with related comorbidities.

References

(1) Prevalence of chronic insomnia in adult patients and its correlation with medical comorbidities: (PMCID - PMC5353813).
(2) Cognitive Behavioral Therapy for Insomnia Comorbid with Psychiatric and Medical Conditions: A Meta-analysis: (PMCID - PMC26147487).
(3) Clustering Insomnia Patterns by Data from Wearable Devices: Algorithm Development and Validation Study: (PMCID – PMC6923760).
(4) Development of Parasomnias from Childhood to Early Adolescence Comorbidities: (PMID - 10878151).
(5) EEG Functional Connectivity Prior to Sleepwalking: Evidence of Interplay between Sleep and Wakefulness: (PMID - 28204773).
(6) Sleep Medicine: Parasomnias: (PMID - 30228711).
(7) Obstructive sleep apnea: current perspectives: (PMCID – PMC5789079).
(8) Deep Analysis of EIT Dataset to Classify Apnea and Non-Apnea Cases in Neonatal Patients: (Art. No. – 9344679/journal-IEEE Access).
(9) Obstructive Sleep Apnea and Risk of Cardiovascular Events and All-Cause Mortality: A Decade-Long Historical Cohort Study: (PMCID – PMC3913558).
(10) Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy | Nature Communications: (PMCID - PMC6283836).
(11) Exploring the clinical features of narcolepsy type 1 versus narcolepsy type 2 from European Narcolepsy Network database with machine learning: (PMCID - PMC6045630).
(12) Supervised and unsupervised machine learning for automated scoring of sleep-wake and cataplexy in a mouse model of narcolepsy: (PMCID - PMC7215268).
(13) Assessment of Cardiac Autonomic Functions by heart rate variability in patients with Restless Leg Syndrome: (PMID - 29664425).
(14) The Effect of Selenium Administration on Restless Leg Syndrome Treatment: (PMID - 22737548).
(15) Assessment and prediction of Restless Leg Syndrome (RLS) in patients with diabetes mellitus type-II though Artificial Intelligence (AI): (PMID - 33832881).
(16) Sleep Disorder Classification Method based on Logistic Regression with Apnea-ECG Dataset: (DOI:10.1145/3358331.3358344).
(17) What is Hypopnea? | Sleep Foundation.
(18) Data | ISRUC-SLEEP Dataset.