

Predictive Analytics for Obstructive Sleep Apnea Detection

M.V.Vijaya Saradhi, Kalli Srinivasa Nageswara Prasad

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. If such comprehensive approach can be developed, it can be effective process for developing a sustainable solution.

Keywords: obstructive sleep apnea, ahi index, machine learning, contemporary information system.

I. INTRODUCTION

Healthcare is one of the critical concerns for global nations. Irrespective of whether it's a developing nation or the super economic nations, in common one of the key areas of focus is the quality healthcare to the public. Alongside the increasing amenities and developments in the healthcare solutions, even the complexities too are on rise. For instance, some of the challenges that are impact the quality health for the individuals is the stress related health problems, obesity related and the contemporary issues like the sleep apnea conditions that are adding up the complexities of quality health for public [1].

Numerous studies are focusing on the issue of Obstructive Sleep Apnea. While the medical studies are aiming at the facets like the root cause analysis, diagnostic patterns and the treatment solutions, some of the IT studies are focusing in the dimension of how the contemporary information systems and the big data kind of analytics can be resourceful in addressing the problem.

1.1 Sleep Apnea

Sleep Apnea is a common phenomenon of sleep disorder, which is faced by millions of people across the countries. The condition is about how a person stops breathing during the sleep, in periodical instances. The cessations of breathing might take place regularly and the intensity and

frequency of cessations decide the categorization as mild or medium or high levels of sleep apnea problem [2].

The impact of the problem is about the brain is not allowed to deeper stages of sleep to prevent choking for breath in the patient. When there are frequent interruptions, the quality of sleep gets affected, thus leading to more health complications.[2]

While there are many who are familiar with the distinct forms of sleep apnea conditions, OSA (Obstructive Sleep Apnea) is the major challenge that goes often unrecognized. The Figure 1 indicates the various types of sleep apnea and its relative symptoms, causes and treatments.

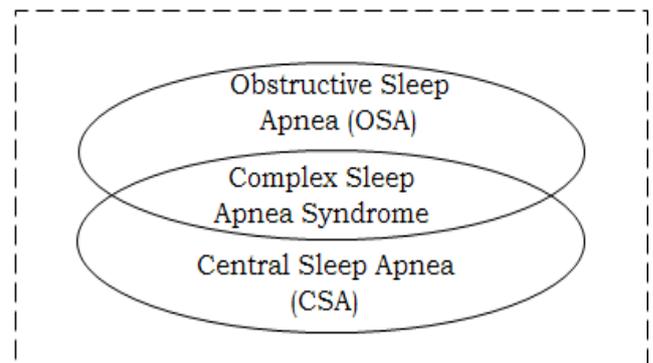


Figure 1: Three Types of Sleep Apnea

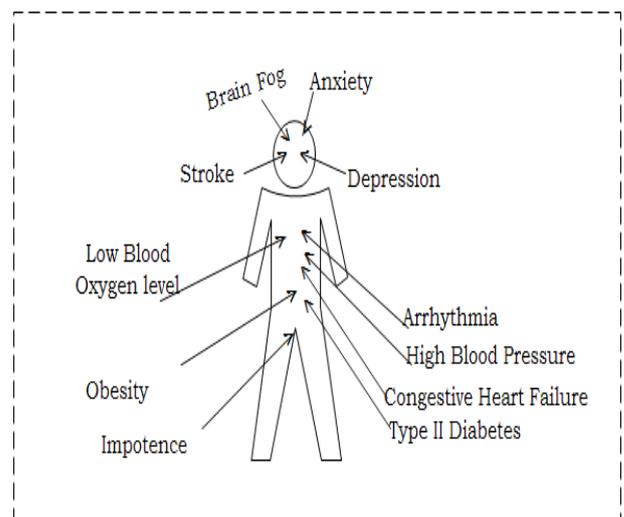


Figure 2: Side Effects from Sleep Apnea

Revised Manuscript Received on July 18, 2019.

Dr.M.V.Vijaya Saradhi, Professor, Dept. Of CSE, ACE Engineering College, Ghatkesar, Hyderabad, India.(Email: meduri.vsd@gmail.co)

Dr. Kalli Srinivasa Nageswara Prasad, Professor, Department CSE, Ramachandra College of Engineering, Eluru, Andhra Pradesh.India. (Email: kallinprasad@gmail.com.)

1.2 Effects of Sleep Apnea

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. Considering the implications that the sleep apnea can lead to the people as discussed in the Figure 2, the need in terms of focusing on the early detection models and the prevention models too are highly important. [3]

Currently, the popular model of diagnosis for the sleep apnea is the sleep study test which the patient undergoes in a clinical set-up. Based on the AHI levels that are resulting in the studies, the presence and the intensity of the sleep apnea is decided for the patients.

The indexing in the current set of sleep apnea diagnosis is on the following set of ranking for AHI index.

- If the AHI index is lower than 5 per hour, then the patient condition is normal and not prone to sleep apnea
- If the AHI index is between 5-14 per hour, classified as a mild sleep apnea prone
- If the AHI index is between 15-29 per hour, then it is moderate sleep apnea conditions
- If the AHI index is between 30 or more instances per hour, then it is considered as severe category of sleep apnea.

Despite the levels of accuracy that is available with the sleep study patterns, still some of the conditions that impact the conditions of conducting the sleep study test could be attributed to the infrastructure set-up and the perceptual issues from the patients. Categorically, in India, though there are many people who are currently suffering sleep apnea or prone, do not have labs available for the regular sleep study tests. There are very limited number of medical institutions or private medical lab establishments that are focusing on the sleep apnea conditions.

Considering the enormity of the conditions, there is integral need for focusing on the alternative models of diagnostic conditions that can support in early detection and the stage detection conditions for the sleep apnea models. In the earlier studies, there are some predictive analytics and other models that were proposed for the sleep apnea detection among the patients. This study focuses on the alternate models that were proposed for sleep apnea conditions, which can help in formulating the scope for future research and solution development in the domain.

II. RELATED WORK

Numerous works have taken place in the domain of sleep apnea conditions in the last two decades of time. There are dozens of studies that has focused over the detection stages of the sleep apnea conditions. Right from the parameters modeling to the detection models, there are multiple factors that were proposed in the studies and has significant results that are integral to the process.

In [4] the authors studied on the conditions wherein the clinical suspicious cases were reviewed by conducting the polysomnography conditional study. In simple terms, the patients who are considered suspicious at the clinical analysis were taken in to the sleep study tests wherein the polysomnography conditions are evaluated in detail. The

sociodemographic conditions and the ESS scale analysis were considered significant in the apnea detection process. BMI, SpO2 and other metrics were measured as integral to the process. The study relied on the multiple regression linear model for detection, wherein the AHI (Apnea Hypopnea Index) was the dependent variable that impacts the conditions.

The results of the study reflect were predominantly based on the regression analysis, wherein the dependent variables chosen for the process are neck circumference, age, allergic rhinitis, and the BMI for the dependent variable condition of AHI. The inputs in the study reiterates on the conditions of how significant the polysomnography conditions are highly important to assimilate the inputs from the clinical prediction conditions. [4]

Profoundly, identifying symptoms in the patients is very important to ascertain the possible impact of sleep apnea in patients. Preventive approach, of relying on the clinical models of symptoms bases suspicion of patients for sleep apnea, which can be recommended for the sleep study can be helpful filtering procedure [5].

In [6], the authors of the study emphasize on the conditions as to how the current set sleep study process can be labor-intensive, inaccessible and is time consuming. As an alternative, the study suggests about the efficiency of ESS (Epworth Sleepiness Scale) and a contemporary prediction model towards evaluating the patients. Based on the results of the test study, the emphasis is that that there is high specificity of around 82.77%, with confidence interval values of around 77.36 in the case of the ESS metric alone.

But when the proposed model of taking in to account the other metric like the age, ESS score, circumference and SaO₂ (minimum oxygen saturation) is taken in to account, the quantum of higher sensitivity, and the confidence interval levels are increasing by higher levels. The study highlights on the conditions wherein the cumulative model's metrics-based solution is performing better than that of the ESS based analysis alone. [4].

In [7], the researchers 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. The study with clinical analysis of the patients emphasize the correlation of OSA to the patients suffering with neck and head pain conditions persistently. However, the conditions also indicate levels of hypertrophied tonsils only in small portion of the patients affected with OSA. But, the nasal obstruction as a condition is very high in the case of the sleep apnea conditions, wherein the anatomical conditions have significant impact [7].

In another research study [8], the scholars are of viewpoint that profoundly the models suggested for OSA detection are reliant on dichotomizing a population and hardly any functional factors were considered as a solution. The possibility of using the STOP-Bang Model (SBM) was proposed towards classification of the patients in to OSA category based on the severity conditions.

However, the test pattern of the solution is again based on the data of the patients who has undergone polysomnography, and the ones those participated in a questionnaire survey which is converted to STOP -Bang equivalent score. The efficacy of the model was discussed based on the correlation values for the tested data based on regression analysis [8].

The other study [9] has focused on the model of integer-based risk scoring system. The participants data was analyzed for four variables like the BMI, BP, Gender and the self-reported score of snoring. Under the regression tests, the ROC (receiver operating characteristics) were taken as the key metric towards deciding on the severity of sleep apnea among the participants. But in relative terms, the conclusion indicates no major difference in the diagnostic performance of the model when compared to the other models proposed in the domain.

Another interesting compilation of the OSA related studies were carried out in [10]. The study indicates on how the conditions of testing of Polysomnography (PSG) as in-home and laboratory test conditions might vary. The study also discussed about the intermediary analysis wherein the coefficient conditions of the models were evaluated using the correlation coefficients or specificity or the sensitivity conditions that prevail in the structure, where the ROC curves are considered. The study reiterates that as there is absence of any specific gold standard conditions, the ROC curves are taken up as minimal values.

A self-structured or home modeled analysis model was discussed in [11] where the demographic and anthropometric data are relied upon for detecting the OSA. ROC analysis was used towards validating the two-channel Apnea link device solution. The test results of the data were evaluated in comparison to the PSG study. The home oximetry process with the support of two-stage model of screening process has significant importance in detecting the conditions that push for the verification of the data.

Area under the curve (AUC) and 2×2 contingency tables were employed to obtain the performance of the new instrument was proposed in [12], where the two-item model-based analysis were carried out, which indicates better outcome. The model is compared to the SBD and the STOP-BANG models discussed in this section, and the results are iterated as similar kind of performance in likes of the models like SBD and the STOP-BANG.

The other distinct model proposed in the solution are about Portable OSA-CAD system which functions based on single channel ECG. The model is about the automated identification of the ECG signal based on optimal two-band FB (Filter bank) techniques. In the designing of the FB, the purpose that researchers chosen is to reduce the spectral localization depending on the accurate regeneration and regularity parameters. The filter bank is used for classifying the wavelet frequency bands in to FUEN (fuzzy entropy) and the LOEN (log of signal-entry) [13]

The learning models like the KNN, DT, Logistic regression, SVM (Support vector machine), linear discriminant was used for learning patterns of the normal and the OSA data sets. With above 90% performance of the model over accuracy, sensitivity and specificity, the model turns to be one of the promising solutions.

In [14], like the ECG model discussed in [13], the other ECG based method was proposed. Unlike earlier model wherein the feature selection is important which is highly dependent on the subject matter expertise and is subjective, the proposed model is a deep neural network and HMM (Hidden Markov Model) that relies on the single-lead ECG signal.

The method depends on the sparse auto-encounter for learning features that depends on the unsupervised learning process. Using the SVM and ANN classifiers, the models are trained for testing. The study iterates on 85% accuracy as the outcome.

The authors of [15] has proposed an automated classification algorithm is proposed that is enabling in better king of performance models which can lead to significant improvements in the conditions of test analysis with improved accuracy and quick turnaround time of developments.

The other model proposed in the literature is the EEC (electroencephalography) which is a single lead system proposed for discriminating the apnea patients and healthy subjects towards managing the complex conditions of classifying apnea and non-apnea patient datasets.

The model proposed is about usage of the unique multiband subframe based feature extraction scheme for focusing on feature variation pattern using statistical measures and probability density functions. The Rician model parameters combined with statistical measure offer significant qualities in terms of standard performance criteria and geometric separability index. The results that are emanated from the experimental studies indicate the reliability of the model. [16].

A study [17] has focused on the contemporary models that were proposed as integral to analysis. Based on the inputs garnered from the four models, the efficacy of the models STOP, STOP-BANG, EPS and four-vector analysis models were compared. The study reflects on the STOP-BANG as a potential model among the four models that are chosen for comparative analysis.

Based on the inputs reviewed in the related work, it is imperative that the studies and the research developments are taking place in the domain, in distinct dimensions. The scope of using the predictive analytics modeling, usage of the machine learning models is turning to be among the key focus area in the study.

III. MACHINE LEARNING IMPLEMENTATION

The contemporary range of dataset classifiers used for the data set training of the machine learning algorithms (supervised and unsupervised learning) models are 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.[18]

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.

In overall, from the review of the literature, one of the common factors across various kind of OSA detection solution models that can make significant difference to the system of PSG based analysis. It is evident that the PSG model can be more robust model of detection but has significant amount of time and resource consumption towards completing the process.

For instance, when there are multiple set of people who are identified as prone for the OSA, by the clinical investigations are recommended for a sleep study test, under the limited set of resources available for the process, the time consumed towards confirming the suspect cases can be time consuming. Hence, there is imperative need for the organization to focus on the conditions that push for better kind of system outcome.[19]

In the other dimension, if the other model of OSA test models can be developed, even in the absence of robust infrastructure, the model can be more significant in terms of limited time spent in the first level detection process. Such conditions play a critical role in ensuring that the models that are significant and enable detection of the OSA cases even in the large-scale models [6].

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 [7].

However, some of the critical factors that are to be taken in to account in the planning and developing of the machine learning models for the OSA solutions are Classification: In the process of ensuring that the accuracy levels are improving significantly, it is paramount importance to understand how the data is being labelled and a class is being assigned. The readings of the data are considered as positive, negative, moderate, mild etc. depending on the various conditions, and it might impact the training dataset quality and further reading models [20].

For instance, in specific to the machine learning models to be applied to the OSA, conditions, following is an illustrative scenario. The blood pressure rate of a human being is one of the metrics that is chosen in the OSA detection analysis. Though there is some standard defined for BP rate by the international medical association, still there can be some +/- to the rate depending on the geographical location where the patient stays, the occupational practices of the patient etc. Now, in the instance, of dataset having one specific set of BP rates as the average, then the scope of results varying in actual to has the possibilities.

Regression is the other key issue that must be considered. Predominantly in the case of the OSA data to be evaluated, the time series values might be modelled for the system. In

such cases the broad range discovery of AHI is the factor that might take place rather than the more accurate conditions. The key point considered here is about getting the exact readings of AHI, even in the absence of the polysomnography tests.

Data clustering is the other major condition to be assess in the system management. There is need for scores of loop-based learning for the system. Though the system is trained by one item dataset, there is imperative need for keeping the system updated with more data from the system evaluation data and from the other set of data that is reviewed by the system for gaining insights. If such a comprehensive support system is developed, it can help the organization in improving the overall quality of the data inputs and the measures that the system could focus upon [21].

3.1 Rule extraction & RESULTS

The critical part in the case of developing a system structure is about managing the rule extraction. Data is relied upon as the fundamental for the extraction of propositional rules. Such rules can be typical and the methods towards discovering the relationships among the data metrics in statistical models, based on the attributes in the data. For instance, in the case of the OSA, the vehicle driving related behavior of the patient in comparison to the sleep index rates of the last few days.

When the conventional data alone is used in the feature selection, there are scope that the initial tendencies for the symptoms of OSA are ignored. The fundamental principle of using the big data in OSA detection process is about the veracity and variety of data that is garnered in high volumes and velocity with which the data is garnered.

Factually, the velocity of the data might not be a right kind of term in the case of the predictive analytics model for OSA. As profoundly the data can be trained as new batch of records for the training set. If there is holistic approach in the system, wherein the data is also added effectively to the training sets regularly, it will help in improving the overall success rate of the systems and the data process [22].

Hence, as an integral characteristic of the predictive modeling and the machine learning structure, the efficacy of the system lies in the quantum of data available for the training datasets and the features that are integral to using them effectively as a training solution.

In the feature selection towards the OSA detection using the machine learning model, some of the key features that are used in the testing data and the score evaluation is [23].

ESS Score	Blood Pressure	BMI
Height	Weight	Age
Headache if any related data	Snoring data	Neck pain related data if any
Waist circumference	Alcoholic behavior	Smoking habits
Obesity conditions	Breathing problems	Other ailments



Table 1: Conventional metrics (Functional metrics)

But in addition to the factors mentioned in the Table 1, there are many other sets of non-conventional inputs too that can be significant for the detection. For instance, the data in the table above is profoundly about the cause factors of OSA which is taken in to feature training process. However, even based on the symptoms too the conditions can be observed. In the initial level clinical examination, the physicians are recommending the condition based on the symptoms that are discussed or traced with the patient.

Some of the symptoms that are leading to sleep apnea conditions are

Loud snoring	decreased interest in sex	gaspings sensation	mood changes
Waking up with a very sore or dry throat	Forgetfulness,	Occasionally waking up with a choking	lack of energy during the day
Morning headaches	Sleepiness while driving	Restless sleep	Recurrent awakenings
Excessive Fatigue	Diabetes	Drowsiness	Insomnia

Table 2: Unconventional metrics (symptoms-based metrics)

Around 16 features that are profound symptoms of the OSA situation in an individual is listed in the Table 2, which is considered significant symptoms towards the OSA. If there is a machine learning based scoring system that is generated and trained with the datasets, the model can be more effective, wherein even the patients can answer the survey questions or readings and have the self-study executed towards understanding the index rate pertaining to the problem. The other critical aspect is to keep the process simple and free of any complexities that plague the implementation of the system.

If such processes are focused upon, there are ways in which the systems can be more emphatic and the methods that are integral to handling the early detection process of OSA can be executed even in the absence of robust infrastructure.

The key advantage that is perceived in the case of symptom-based feature selection model is the ease at which the patient can answer the questions pertaining to the symptoms. However, if the solution can be a combination of multi-layered analysis wherein the first stage can be symptom-based and the second layer as functional parameters based, it can lead to more accurate model of detection systems.

Considering the success rate of STOP-BANG or STOP or the four vector models, if the feature selections are effective and the data is evaluated using any such models, the scope of accuracy in the system can be significantly improved. With greater number of patients being recommended for the PSG tests, if such unconventional and robust models are developed, it can turn to be a sustainable solution.

IV. 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.

V. REFERENCES

1. Mencar, Corrado, et al. "Application of machine learning to predict obstructive sleep apnea syndrome severity." *Health informatics journal* (2019): 1460458218824725.
2. Bozkurt, Selen, Asli Bostanci, and Murat Turhan. "Can Statistical Machine Learning Algorithms Help for Classification of Obstructive Sleep Apnea Severity to Optimal Utilization of Polysomnography Resources?" *Methods of information in medicine* 56.04 (2017): 308-318.
3. Stretch, Robert, et al. "0465 Performance of Revised Machine Learning Models for Prediction of Non-Diagnostic Home Sleep Apnea Tests." *Sleep* 42. Supplement_1 (2019): A187-A187.
4. Musman, Silvio, et al. "Evaluation of a prediction model for sleep apnea in patients submitted to polysomnography." *Journal Brasileiro de Pneumologia* 37.1 (2011): 75-84.
5. Myers, Kathryn A., Marko Mrkobrada, and David L. Simel. "Does this patient have obstructive sleep apnea? The Rational Clinical Examination systematic review." *Jama* 310.7 (2013): 731-741.
6. Zou, Jianyin, et al. "An effective model for screening obstructive sleep apnea: a large-scale diagnostic study." *PLoS one* 8.12 (2013): e80704.
7. Zonato, Adriane I., et al. "Association of systematic head and neck physical examination with severity of obstructive sleep apnea—Hypopnea syndrome." *The Laryngoscope* 113.6 (2003): 973-980.
8. Farney, Robert J., et al. "The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index." *Journal of Clinical Sleep Medicine* 7.05 (2011): 459-465.



9. Takegami, Misa, et al. "Simple four-variable screening tool for identification of patients with sleep-disordered breathing." *Sleep* 32.7 (2009): 939-948.
10. Secretariat, Medical Advisory. "Polysomnography in patients with obstructive sleep apnea: an evidence-based analysis." Ontario health technology assessment series 6.13 (2006): 1.
11. Chai-Coetzer, Ching Li, et al. "A simplified model of screening questionnaire and home monitoring for obstructive sleep apnoea in primary care." *Thorax* 66.3 (2011): 213-219.
12. Duarte, Ricardo LM, et al. "Simplifying the screening of obstructive sleep apnea with a 2-item model, No-Apnea: a cross-sectional study." *Journal of Clinical Sleep Medicine* 14.07 (2018): 1097-1107.
13. Sharma, Manish, Mitesh Raval, and U. Rajendra Acharya. "A new approach to identify obstructive sleep apnea using an optimal orthogonal wavelet filter bank with ECG signals." *Informatics in Medicine Unlocked* (2019): 100170.
14. Li, Kunyang, et al. "A method to detect sleep apnea based on deep neural network and hidden markov model using single-lead ECG signal." *Neurocomputing* 294 (2018): 94-101.
15. Almazaydeh, Laiali, Khaled Elleithy, and Miad Faezipour. "Detection of obstructive sleep apnea through ECG signal features." 2012 IEEE International Conference on Electro/Information Technology. IEEE, 2012.
16. Bhattacharjee, Arnab, et al. "Sleep Apnea Detection Based on Rician Modeling of Feature Variation in Multiband EEG Signal." *IEEE journal of biomedical and health informatics* 23.3 (2018): 1066-1074.
17. Silva, Graciela E., et al. "Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales." *Journal of Clinical Sleep Medicine* 7.05 (2011): 467-472.
18. Huang, W., et al. "0495 Prediction of Obstructive Sleep Apnea Using Machine Learning Technique." *Sleep* 41 (2018): A186.
19. Pomprapa, A., et al. "Classification of Obstructive Sleep Apnea Using Machine Learning." D102. Improving Identification, Adherence, and Outcomes In Sleep Disordered Breathing. American Thoracic Society, 2018. A7449-A7449.
20. <https://healthitanalytics.com/news/deep-learning-tool-matches-human-experts-at-sleep-test-scoring>.
21. Chu, Kuo-Chung, Hsin-Jou Huang, and Yu-Shu Huang. "Machine learning approach for distinction of ADHD and OSA." 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, 2016.
22. Urtnasan, Erdenebayar, et al. "Optimal classifier for detection of obstructive sleep apnea using a heartbeat signal." *International Journal of Fuzzy Logic and Intelligent Systems* 17.2 (2017): 76-81.
23. Mencar, Corrado, et al. "Application of machine learning to predict obstructive sleep apnea syndrome severity." *Health informatics journal* (2019): 1460458218824725