Biomarker feature selection for detecting Obstructive Sleep Apnea events

PIPER STACEY, Dartmouth College
VAFA BATOOL, Dartmouth College
JOHN BERRY, Dartmouth College
SHITONG CHENG, Dartmouth College

The abstract is here!

Additional Key Words and Phrases: datasets, sleep apnea, machine learning, text tagging

ACM Reference Format:
Piper Stacey, Vafa Batool, John Berry, and Shitong Cheng. 2022. Biomarker feature selection for detecting Obstructive Sleep Apnea events . 1, 1 (October 2022), 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Obstructive sleep apnea is a form of disordered breathing while sleeping that occurs particularly due to an obstruction in the upper airway [2]. An estimated 80% to 90% of all obstructive sleep apnea cases go undiagnosed [15]. The problem with obstructive sleep apnea syndrome is that patients do not get enough high quality sleep. Lower quality sleep leads to accidents while higher quality sleep promotes attentiveness and reduces blood pressure. Furthermore, there is a known association between obstructive sleep apnea and hypertension, metabolic syndrome, diabetes, heart failure, coronary artery disease, arrhythmias, stroke, pulmonary hypertension, neurocognitive and mood disorders [10]. Specifically in stroke patients, untreated sleep apnea is associated with major risk factors for another stroke. The prevalence of sleep apnea is 50%-70% in stroke patients. Earlier diagnoses and treatments of sleep apnea in stroke patients could improve recovery from stroke and reduce the likelihood of another stroke in the future [2].

There has been a lot of work studying ways to leverage data science to detect sleep apnea but they have focused on only a few datapoints across the range of what is commonly collected. There has not yet been a study that seeks to determine which of those commonly focused on result in improved accuracy when automatically detecting sleep apnea events.

1.1 Objective and Search Plan

We have many dimensions of data, such as HR, SpO2, RR, etc, not every type of data is sensitive to the detection of apnea, or they are well suited as data for detecting apnea. It is also known that the choice of model or algorithm will determine the accuracy of the final result to some extent. We strive to select the optimal combination of subsets of the data and algorithmic, such that the detection of apnea is as accurate and sensitive as possible.

Authors' addresses: Piper Stacey, piper.f.stacey.23@dartmouth.edu, Dartmouth College; Vafa Batool, vafa.batool.gr@dartmouth.edu, Dartmouth College; John Berry, john.n.berry.gr@dartmouth.edu, Dartmouth College; Shitong Cheng, shitong.cheng.gr@dartmouth.edu, Dartmouth College.

XXXX-XXXX/2022/10-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Previous work has been done based upon a variety of learning algorithms though nothing has been published comparing the efficacy of one algorithm over the other in dealing with this problem space. We will select several algorithms, likely some subset of K-means clustering, SVMs, Reinforcement learning, Random forests, and others, upon which to base our modelling. With each of these algorithms we will train models on subsets of the features included in our data set. For example, one model uses HRV, RR, SpO2 and some Polysomnograpy readings while another model uses just HRV and RR. etc. From here, the goal is to determine which combination of models and features produces the greatest performance in respect to accuracy and sensitivity. (Type I and Type II errors).

1.2 Contribution

There has not been a concerted effort to determine if combinations of these data points and algorithms can perform better than the models trained on the data points individually. Using the features provided in [1] such as the perfusion index, respiratory rate, detected snores, and others.This paper presents a comparison between modelling methodologies in conjunction with feature set selections that when combined provide the best predictive capabilities of OSA events in patients. OSA has been associated with many negative health outcomes though early detection and management can reduce these. Finding a model that best detects these events can lead to improved patient treatment plans.

2 RELATED WORK

Previous works in the space of Obstructive Sleep Apnea (OSA) have explored its impact on the quality of life of individual. It includes how it may negatively impact the posture and gait and have attributed it as among a leading cause of falls in older individuals. They have focused on treating OSA and measuring its potential effect, if any, on injuries sustained due to falls [16]. Muraki et. al have also contributed towards understanding how OSA may be correlated with other health conditions like type II diabetes. They help uncover how continuous positive airway pressure (CPAP), the main treatment for OSA, may help improve insulin resistance as a way for better treating diabetes patients [11]. Some past studies have helped uncover how OSA may be contributing to comorbidities like hypertension. Based on the meta-analysis of PubMed and Embase databases, it was found that OSA is linked to increased hypertension, especially in male Caucasian individuals [6]. The relationship between OSA and strokes has also been studied previously. Dyken et. al found a close causal relationship between the two where the events of snoring and recorded instances of apnea before a stroke have revealed that untreated OSA may lead to strokes [3, 15]. Studies that have focused on stroke patients with apnea have also investigated through a longitudinal study the outcome of OSA on their life. It was found that after a year the survival rate of apnea patients, especially men, was significantly lower than the patients without apnea problems [8].

In previous studies, machine learning techniques have proven to be useful with regard to picking bio-markers that are able to predict a patient's future condition [18]. Specifically, researchers have used single-channel nasal pressure airflow signals to diagnose OSA [5]. Other study diagnosed OSA based on hyperparameters such as blood reports, demographics, physical measurements, comorbidities, and sleep habits [13]. Others have used ECG spectrograms to diagnose OSA [4, 9, 17] while some have chosen to focus on using respiratory rate and oxygen saturation [14]. Furthermore, studies have used machine learning and support vector machine predictions on polysomnography data to predict OSA [7, 12]. Finally, in the past there have been attempts at identifying genes associated with OSA and CPAP to improve the diagnostic accuracy [1].

As such, previous works have focused largely on understanding the link between OSA and other health condition, how it can be treated more effectively and its detection based on genes involved.

There has been no attempt at using the biomarkers available such as oxygen levels overnight or respiratory rate to predict the onset of apnea events. This is where the novelty of our work would step in.

REFERENCES

[1] Bernardini, A., Brunello, A., Gigli, G. L., Montanari, A., and Saccomanno, N. Osasud: A dataset of stroke unit recordings for the detection of obstructive sleep apnea syndrome. *Scientific Data 9*, 1 (2022), 1–10.

[2] Davis, A. P., Billings, M. E., Longstreth, W., and Khot, S. P. Early diagnosis and treatment of obstructive sleep apnea after stroke: Are we neglecting a modifiable stroke risk factor? *Neurology: Clinical Practice 3*, 3 (2013), 192–201.

[3] Dyken, M. E., and Im, K. B. Obstructive sleep apnea and stroke. *Chest 136*, 6 (2009), 1668–1677.

[4] Erdenebayar, U., Kim, Y. J., Park, J.-U., Joo, E. Y., and Lee, K.-J. Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. *Computer methods and programs in biomedicine 180* (2019), 105001.

[5] Haidar, R., Koprinska, I., and Jeffries, B. Sleep apnea event detection from nasal airflow using convolutional neural networks. In *International Conference on Neural Information Processing* (2017), Springer, pp. 819–827.

[6] Hou, H., Zhao, Y., Yu, W., Dong, H., Xue, X., Ding, J., Xing, W., and Wang, W. Association of obstructive sleep apnea with hypertension: a systematic review and meta-analysis. *Journal of global health 8*, 1 (2018).

[7] Huang, W.-C., Lee, P.-L., Liu, Y.-T., Chiang, A. A., and Lai, F. Support vector machine prediction of obstructive sleep apnea in a large-scale chinese clinical sample. *Sleep 43*, 7 (2020), zsz295.

[8] Kojić, B., Dostović, Z., Ć Ibrahimagić, O., Smajlović, D., and Iljazović, A. Does sleep apnea affect the outcome of stroke patients? *Psychiatria Danubina 33*, suppl 4 (2021), 496–502.

[9] Lin, C.-Y., Wang, Y.-W., Setiawan, F., Trang, N. T. H., and Lin, C.-W. Sleep apnea classification algorithm development using a machine-learning framework and bag-of-features derived from electrocardiogram spectrograms. *Journal of Clinical Medicine 11*, 1 (2021), 192.

[10] Mannarino, M. R., Di Filippo, F., and Pirro, M. Obstructive sleep apnea syndrome. *European journal of internal medicine 23*, 7 (2012), 586–593.

[11] Muraki, I., Wada, H., and Tanigawa, T. Sleep apnea and type 2 diabetes. *Journal of diabetes investigation 9*, 5 (2018), 991–997.

[12] Palotti, J., Mall, R., Aupetit, M., Rueschman, M., Singh, M., Sathyanarayana, A., Taheri, S., and Fernandez-Luque, L. Benchmark on a large cohort for sleep-wake classification with machine learning techniques. *NPJ digital medicine 2*, 1 (2019), 1–9.

[13] Ramesh, J., Keeran, N., Sagahyroon, A., and Aloul, F. Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning. In *Healthcare* (2021), vol. 9, MDPI, p. 1450.

[14] Ravelo-García, A. G., Kraemer, J. F., Navarro-Mesa, J. L., Hernández-Pérez, E., Navarro-Esteva, J., Juliá-Serdá, G., Penzel, T., and Wessel, N. Oxygen saturation and rr intervals feature selection for sleep apnea detection. *Entropy 17*, 5 (2015), 2932–2957.

[15] Silverberg, D. F., Iaina, A., and Oksenberg, A. Treating obstructive sleep apnea improves essential hypertension and quality of life. *American family physician 65*, 2 (2002), 229.

[16] Stevens, D., Jackson, B., Carberry, J., McLoughlin, J., Barr, C., Mukherjee, S., Oh, A., McEvoy, R. D., Crotty, M., and Vakulin, A. The impact of obstructive sleep apnea on balance, gait, and falls risk: a narrative review of the literature. *The Journals of Gerontology: Series A 75*, 12 (2020), 2450–2460.

[17] Tison, G. H., Zhang, J., Delling, F. N., and Deo, R. C. Automated and interpretable patient ecg profiles for disease detection, tracking, and discovery. *Circulation: Cardiovascular Quality and Outcomes 12*, 9 (2019), e005289.

[18] Zhang, X., Jonassen, I., and Goksøyr, A. Machine learning approaches for biomarker discovery using gene expression data. *Exon Publications* (2021), 53–64.