ECG, EEG, Breathing Signals, and Machine Learning: Computer-Aided Detection of OSAS and Depression in OSAS Patients

Mostafa M. Moussa, Yahya Alzaabi, and Ahsan Khandoker, *Senior Member, IEEE*

***Abstract*— Obstructive Sleep Apnea Syndrome (OSAS) and Major Depressive Disorder (MDD) are both common conditions associated with poor quality of life. In this work, we seek to classify OSAS and depression in OSAS patients using multiple machine learning techniques. We have extracted features from 5-minute intervals selected from electrocardiograms (ECG), breathing signals, and electroencephalograms (EEG) recorded from a total of 118 subjects, of which 89 are used for training and 10-fold cross-validation and 29 are used for testing or a 75/25% split. This dataset was used as the input to three classification problems: sleep staging, classifying OSAS, and depression in OSAS patients. The best classification performance of OSAS was obtained with light sleep and deep sleep with the ReliefF feature selection algorithm using random forest and boosted trees for classification, respectively. It has yielded an accuracy of 100.00%, F1-Score of 100.00%, Cohen's κ Coefficient of 1.00, and a Matthews correlation coefficient (MCC) of 1.00. All sleep stages with 10 principal components following principal component analysis (PCA) and using random forest for classification yielded an accuracy of 77.50%, F1-Score of 78.05%, Cohen's κ of 0.571, and an MCC of 0.632 for classification of depression in OSAS patients. Sleep staging was best done using bagged trees with features selected via sequential backward feature selection, yielding an accuracy of 76.90%, F1-Score of 75.90%, Cohen's κ of 0.480 and an MCC of 0.634. These results show promise in detecting OSAS and depression in OSAS patients, particularly using labeled sleep stage data, light, and deep sleep.**

***Index Terms*— Obstructive Sleep Apnea Syndrome (OSAS), Depression, Sleep staging, Machine learning.**

I. Introduction

OBSTRUCTIVE Sleep Apnea Syndrome (OSAS) is a prevalent condition with 3-7% of men and 2-5% of women in the general population [1] suffering from it with an associated decrease in sleep quality and overall quality of life, due to its many comorbidities. These include strokes, coronary heart disease (CHD), mood disorders, memory loss, and reduced cognitive performance [2–4]. Electrocardiography (ECG) is among the most common electrophysiological signal recording techniques due to the importance of reading heart signals for countless applications and tests, and as such, it is also commonly recorded in sleep studies, or polysomnography (PSG), often for use with machine learning algorithms.

Works like Khandoker *et al.*'s, Bozkurt *et al.*'s and Erdenebayar *et al.*'s focus on the use of support vector machines (SVMs), multiple classic classifiers, and convolutional and recurrent neural networks (CNNs and RNNs), respectively [4–6]. The first of which obtained a testing accuracy of 92.85% and a Cohen's κ coefficient of 0.85 with 83 out of 125 subjects after training SVMs with heart rate variability (HRV) and ECG-derived respiration (EDR) features and leave-one-out cross-validation [4]. Bozkurt *et al.* derived HRV and QRS intervals from 10 ECGs, and in turn, extracted features and performed feature selection using Fisher's algorithm. An ensemble of decision trees (DT), Kth-nearest neighbor (KNN), and SVM and 50% of the features selected, yielded the best performance with an accuracy of 85.12%, sensitivity of 85.00% and specificity of 86.00% [5]. Finally, Erdenebayar *et al.* use gated recurrent unit (GRU) neural networks, which are a type of long short-term memory (LSTM) recurrent neural networks (RNN) with 80% of 86 subjects used for training and the remainder for testing, to classification accuracy and sensitivity of 99.00% [6].

Works regarding the detection of OSAS and depression are somewhat rare, though machine learning is a relatively standard methodology for detecting these individual conditions. Furthermore, the discussed works commonly extensively preprocess data in order to have robust features for machine learning. Hence, the contribution of this work centers around the idea of using features extracted from 5-minute intervals from the ECG, EEG, and breathing signals following filtering and re-sampling as necessary, selecting features among these using various algorithms. Afterward, we use these features to train machine learning algorithms to perform sleep staging, classify OSAS, and classify depression in OSAS patients.

We describe the methodology utilized in Section II and the results obtained from following said methodology in Section III, we discuss our results in Section IV. Finally, we end with concluding remarks and possible avenues for future work in Section V.

This work was funded by the KING ABDULAZIZ UNIVERSITY-KHALIFA UNIVERSITY JOINT RESEARCH PROGRAM (KAU-KU JRP), KU Project number 8474000376.

Submitted August 2022.

Mostafa M. Moussa is with the Department of Biomedical Engineering, Khalifa University of Science and Technology, Abu Dhabi, UAE, 127788. (e-mail: mostafa.moussa@ku.ac.ae).

Yahya Alzaaabi is with the Department of Biomedical Engineering, Khalifa University of Science and Technology, Abu Dhabi, UAE, 127788. (e-mail: yahya.alzaabi@ku.ac.ae).

Ahsan Khandoker is with the Department of Biomedical Engineering, Khalifa University of Science and Technology, Abu Dhabi, UAE, 127788. (e-mail: ahsan.khandoker@ku.ac.ae).

II. Methodology

Overnight polysomnography was performed on 86 subjects at the American Center for Psychiatry and Neurology, among which 40 had OSAS alone, 40 had OSAS paired with depression. In addition, 32 subjects without OSAS or depression are taken from the Stanford Technology Analytics and Genomics in Sleep (STAGES) dataset [7] to form a total of 118 subject-dataset with 59 male and 57 female subjects, of which 89 or 75% are used for training and 10-fold cross-validation and the remaining 29 or 25% are used for testing. The methodology is briefly described in Figure 1.

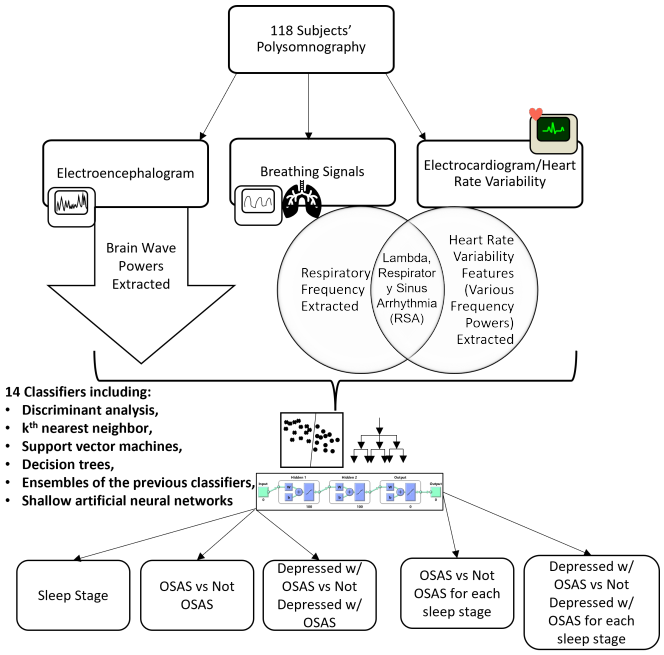

Fig. 1. Methodology for sleep staging, classification of OSAS, and classification of depression in OSAS patients.

We mainly focus on ECG, EEG, and breathing signals (flow and thorax signals), the first two were sampled at 100 Hz and 200 Hz, respectively, and were re-sampled to breathing signals' sampling frequency, 10 Hz, to synchronize the signals in time prior to feature extraction. Power-line interference is filtered out of the electrophysiological signals via a 50 Hz notch filter, followed by band-pass filtering as appropriate before re-sampling. Finally, the signals are split into 5-minute intervals, from which features are extracted based on the status of depression, apnea, and sleep stage during that interval, with depression being dependent on the subject themselves, rather than the interval.

Classification of OSAS and classification of depression in OSAS patients are done both with all data, and the data for each individual sleep stage, while sleep staging is logically only done with all data. Since we begin by classifying whether or not the subjects have OSAS using subjects with OSAS regardless of depression status vs the control group, the control group is removed afterward when we seek to classify depression. This means that the number of intervals or observations available for the classification of depression is less than the number of observations available for detection of OSAS and sleep staging (with all data only). Consequently, the selected features for the classification of depression would also differ from those selected for the classification of OSAS and for sleep staging.

The features extracted include respiratory frequency extracted from breathing signals, and brain wave powers extracted from EEG channels O1, O2, C3, C4, F3, and F4 with A1 and A2 (or M1 and M2) as reference channels. Furthermore, very low-frequency, low-frequency, and high-frequency powers, as well as the ratio of low-to-high-frequency powers of ECG signals or heart rate variability signals, and features pertaining to both ECG and breathing were extracted such as lambda, which represents the phase coupling between R-R intervals in the ECG and thorax signal, and respiratory sinus arrhythmia (RSA), which represents the reduction in R-R interval duration during inspiration and extension during expiration. The resultant features are thus lambda, RSA, respiratory frequency, normalized RSA, vLF power, LF power, normalized LF power, HF power, normalized HF power, LF/HF, and the band powers of beta, theta, alpha, and delta from each of the 6 EEG channels aside from the reference ones for a total of 34 features.

No further processing is applied to this data, but features are selected using one of 7 techniques/algorithms: sequential forward feature selection (SFFS), sequential backward feature selection (SBFS), minimum redundancy maximum relevance (MRMR), ReliefF [8], neighborhood component analysis (NCA), Chi-Squared, principal component analysis taking the first 10 components, and principal component analysis taking components that explain 95% of the variance.

After feature selection, the data is ready to be input into our classifiers. We essentially have two sets of problems, the first set involves three problems: sleep staging, done in parallel with the classification of OSAS, and classification of depression in OSAS patients. The second set involves the latter two problems performed with data from each individual sleep stage. This means we have, overall, three classification problems, the first of which is 3-class due to the way sleep stages were denoted (light sleep, REM sleep, and deep sleep were considered; light sleep is not split into two stages), and the second and third are binary. The number of classes in our problems, however, does not restrict our use of any classifiers, as we end up using up to 14 classifiers including Gaussian Distribution Naive Bayes (NB), linear discriminant analysis (LDA) with a regularization term γ of 0, decision tree (DC) with 100 maximum number of branch nodes split by Gini's diversity index (GDI), Euclidean distance k^{th} nearest neighbor (KNN) with k set to 1, radial basis function (RBF) support vector machine (SVM) with automatically computed kernel scale and a box constraint of 1, bagged trees with 1290 maximum number of branch nodes and 30 learning cycles, random forest (RF), boosted trees with 4 algorithms (AdaBoost, RUSBoost, LogitBoost, GentleBoost) and a maximum number of branch nodes of 30, subspace KNN and discriminant with 30 learning cycles, and a 4-layer artificial neural network (ANN), wherein there are an input and output layers and two hidden layers with 100 units in each.

III. Results

In order to measure the performance of our classifiers, we use several well-known metrics in machine learning, namely accuracy, sensitivity (also known as recall), specificity, precision, F1-score, Cohen's Kappa (κ) coefficient, Matthews correlation coefficient (MCC), and the area under the receiver operating characteristics (ROC) curve, also known as AUC. These metrics are described in Equations 1 through 7, where TP, FP, FN, and TN are true positive, false positive, false negative, and true negative instances and the expected accuracy is the accuracy when the instances are classified by chance, based on the number of classes and number of instances in each class.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{FP + TN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1 - Score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \tag{5}$$

$$= \frac{Accuracy - ExpectedAccuracy}{1 - ExpectedAccuracy} \tag{6}$$

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

Sleep Staging is best done with SBFS using bagged trees for classification with an AUC of 0.846, an accuracy of 76.90%, sensitivity of 75.38%, specificity of 87.04%, precision of 77.03%, F1-score of 75.90%, Cohen's κ of 0.480, and a Matthews correlation coefficient of 0.634. OSAS is best classified with ReliefF using light and deep sleep data, using random forest and boosted trees with GentleBoost algorithm, respectively for classification. These algorithms with their respective data and feature selection techniques yielded an AUC, Cohen's $kappa$, and Matthews correlation coefficient of 1.00, as well as accuracy, sensitivity, specificity, precision, and F1-score of 100%. Depression in OSAS patients is best classified with PCA taking the first 10 PCs using deep sleep data, using a random forest for classification with an AUC of 0.880, an accuracy of 77.50%, sensitivity of 64.00%, specificity of 100.00%, precision of 100.00%, F1-score of 78.05%, Cohen's κ of 0.571, and a Matthews correlation coefficient of 0.632. These results are summed up in Table I, and Figure 2 provides a visual representation in the form of posterior probability and box plots.

TABLE I
BEST CLASSIFICATION RESULTS FOR EACH PROBLEM.

	Sleep Staging	**Classification of OSAS**		**Classification of Depression in OSAS Patients**
Sleep Stage	**N/A or All Stages**	**Light Sleep**	**Deep Sleep**	**Deep Sleep**
Feature Selection Technique	SBFS	ReliefF	ReliefF	PCA-10 PCs
Model	Bagged Trees	Random Forest	GentleBoost Boosted Trees	Random Forest
AUC	0.846	1.00	1.00	0.880
Accuracy (%)	76.90	100.00	100.00	77.50
Sensitivity (%)	75.38	100.00	100.00	64.00
Specificity (%)	87.04	100.00	100.00	100.00
Precision (%)	77.03	100.00	100.00	100.00
F1-Score (%)	75.90	100.00	100.00	78.05
κ	0.480	1.00	1.00	0.571
MCC	0.634	1.00	1.00	0.632

IV. Discussion

The computed metrics in Table I show excellent performance for classification of OSAS with these datasets and methodology, as well as decent performance for sleep staging and classification of depression in OSAS patients. Deep sleep seems to be the common denominator in the best classification performance for both OSAS and depression in OSAS patients, allowing us to recommend recording polysomnography signals during deep sleep. Since our classification of OSAS and depression using all sleep stage data was done in parallel with sleep staging, we refrain from making hard recommendations for automation of sleep staging before moving on to classification of OSAS and depression since its classification performance can likely be improved with further processing of the signals or more advanced classification techniques. Moreover, note that the performance using REM sleep data for classification does not appear in our table, which indicates that increased brain activity may be a slight hindrance for all three classification problems when using ECG, EEG, and breathing signals, perhaps because 24 out of the initial 34 features are extracted from the EEG. In addition, the results obtained for OSAS and depression classification using all sleep stage data compared to individual sleep stages is slightly worse, so we recommend clinically or mathematically defining sleep stages prior to classification of OSAS and depression in OSAS patients.

Besides these clinical/technical recommendations, we also notice that decision tree ensemble classifiers, namely bagged trees, boosted trees, and random forest yield the best performance in all three classification problems. The exclusive use of these classifiers would save time and computational cost if the methodology were implemented for the detection of OSAS and depression in OSAS patients in real-life, following data acquisition and feature extraction of deep sleep data intervals.

Future directions in this research should include performing classification using the PSG signals directly without manual feature extraction, which allows better automated analysis of PSG data. This is because feature extraction is not a one-size-fits-all solution, whereas deep learning, for example, is adaptive and perceptive to both intra and inter-subject variability, if trained appropriately. Nevertheless, other methods of improving performance while making minimal changes to the current methodology involve using different extracted features, further pre-processing and processing, increasing the number of subjects, and ensuring minimal subject variability.

A different research avenue that this methodology can be amended and used for involves predicting OSA events before they occur, by training machine learning classifiers using PSG intervals during apnea events, but with more focus on the signals before and after them.