

# Investigating Feature Importance in Sleep Apnea and Hypopnea Detection: A Random Forest and SHAP-based Approach

1<sup>st</sup> Shayan Sharifi

*Dept. of Information Engineering (DEI)*  
*University of Padova*  
Italy

shayan.sharifi@studenti.unipd.it

2<sup>nd</sup> Shahla Sadeghzadeh

*Dept. of Information Engineering (DEI)*  
*University of Padova*  
Italy

shahla.sadeghzadehdarandash  
@studenti.unipd.it

3<sup>rd</sup> Giulia Cisotto

*Dept. of Information Engineering (DEI)*  
*University of Padova*  
Italy

giulia.cisotto@unimib.it

4<sup>th</sup> Alberto Zancanar

*Dept. of Information Engineering (DEI)*  
*University of Padova*  
Italy

alberto.zancanaro.1@phd.unipd.it

**Abstract**—Sleep apnea and hypopnea are prevalent sleep disorders with significant health implications. Accurately identifying these disorders is essential for effective diagnosis and treatment. In this study, we propose a new approach to pinpoint the specific features of each signal that have the greatest impact on detecting sleep apnea and hypopnea. By using a random forest algorithm and SHAP analysis, we examined the St. Vincent's University Hospital dataset, which includes overnight PSG and ECG signals from patients suspected of having sleep-disordered breathing. We extracted time and frequency domain features from the signals, capturing diverse characteristics that indicate sleep apnea and hypopnea. These features were fed into the random forest model. Subsequently, SHAP analysis was employed to unveil the most influential features in the recognition process. The identified key features offer valuable insights into the underlying physiological characteristics and mechanisms associated with sleep apnea and hypopnea. This knowledge opens avenues for improved diagnostic approaches and personalized treatment strategies. Our study underscores the potential of combining machine learning algorithms with interpretability techniques to advance our understanding of sleep disorders.

**Index Terms**—Machine learning; Feature extraction; Random forest classifier; Shapley value Sleep Apnea; Hypopnea;

## I. INTRODUCTION

Obstructive sleep apnea syndrome (OSAS) is a very common disorder with incidence estimated at 5–14 percent among adults aged 30–70 years. The clinical importance of OSAS is related to increased risk of cardiovascular diseases as well as higher morbidity and mortality [1]. The gold standard for diagnosis of OSAS is the polysomnography (PSG) test [2] through which it is possible to have information on the severity

of OSAS and the degree of sleep fragmentation. However, PSG requires overnight evaluation in a sleep laboratory, dedicated systems and attending personnel. [1] Recently, the field of Medicine has embraced innovative data analysis techniques, such as artificial neural networks, to effectively analyze vast volumes of clinical data. This utilization aims to deepen our understanding of the underlying disease mechanisms and enhance diagnostic capabilities for various medical conditions. In this article, we employ a supervised machine learning approach, specifically a random forest classifier [3], to address the classification of apneas, and hypopneas status into six distinct categories: obstructive, central, mixed [4]. Essentially, our classifier assigns one of six statuses, HYP-0/C/M or APNEA-0/C/M, to each patient based on start time and duration of respiratory events. Then we calculate the Shapley value for each signal to determine their respective contributions to the final diagnosis. This allows us to identify the specific features of each signal that have the greatest impact on sleep apnea and hypopnea. These computations are performed through SHAP (Shapley Additive explanations), a software tool to explain the output of any machine learning model through the classic Shapley value from game theory [5].

## II. DATASET

St. Vincent's University Hospital Sleep Apnea Database contains 25 full overnight polysomnograms with simultaneous three-channel Holter ECG, from adult subjects with suspected sleep-disordered breathing [dataset]. Polysomnography (PSG) is the gold standard test for sleep disorder diagnosis [2] and it records multiple channels, it were obtained using the

Jaeger-Toennies system. Signals recorded were: EEG (C3-A2), EEG (C4-A1), left EOG, right EOG, submental EMG, ECG (modified lead V2), oro-nasal airflow (thermistor), ribcage movements, abdomen movements (uncalibrated strain gauges), oxygen saturation (finger pulse oximeter), snoring (tracheal microphone) and body position. Three-channel Holter ECGs (V5, CC5, V5R) were recorded using a Reynolds Lifecard CF system (Reynolds Medical, UK) [4].

### III. PIPELINE

The entire pipeline of our approach is illustrated in “Fig. 2”.

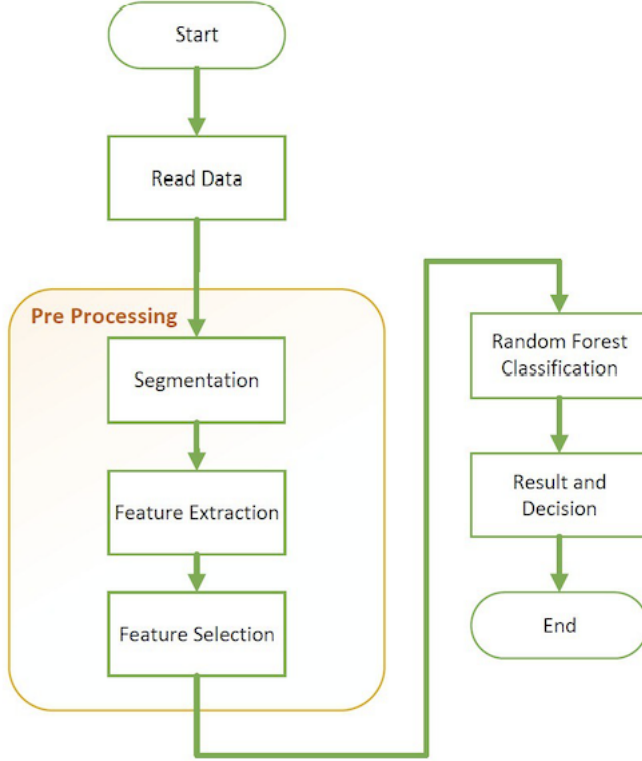


Fig. 1. Pipeline

Our pipeline consists of four key steps to analyze sleep apnea and hypopnea:

- **Data Acquisition:** We start by reading the PSG and ECG signals from the dataset.
- **Preprocessing:** This step involves several subtasks to prepare the data for classification. Firstly, we perform segmentation, where the signals are divided into segments based on the start time and respiratory status. Secondly, we extract relevant features from the segmented signals. Lastly, we employ a feature selection technique, such as the SelectKBest method, to identify the most informative features for classification.
- **Classification:** We utilize the random forest algorithm for classification. The selected features from the previous step serve as inputs to the random forest classifier, enabling it to learn patterns and make predictions based on the training data.

- **Interpretation:** To gain insights into the importance of different features in the classification process, we calculate the Shapley values for the trained random forest model. The Shapley values quantify the contribution of each feature towards the final prediction.

### IV. PREPROCESSING

The primary goal of preprocessing is the selection of a suitable subset of features that will be used to design the predictive models [1]. To achieve this objective, we adopt a two-step approach. Firstly, we perform segmentation, which involves extracting relevant signals based on the corresponding data labels. This step ensures that we isolate the appropriate segments for further analysis. Secondly, we conduct feature extraction, wherein we extract meaningful features from each signal. By extracting and analyzing these features, we aim to uncover the key characteristics that contribute to the classification and understanding of the data.

#### A. Segmentation

Initially, PSG and ECG signals are read from the dataset. Subsequently, a normalization function is applied to each signal, taking into account the amplitude and mean values. This step ensures that the signals are brought to a consistent scale for further analysis.

To organize the information related to the classification label, signal index, and segmented signal values for each patient, we construct a data frame. This data frame is created by extracting six states for apnea/hypopnea based on their corresponding event times. The table below illustrates the structure of the new raw data frame:

TABLE I  
DATA FRAME

Index	Event Type	Signal Index	Values
0	HYP-C	0	[0.02028, 0.02028, ...
1	HYP-M	1	[0.00837, -0.04883, ...
2	APNEA-C	2	[-0.74050, 0.52054, ...
3	APNEA-O	3	[0.76853, 0.77041, ...
...	...	...	...

In this data frame, each row represents a specific segment of the signal corresponding to an apnea or hypopnea event. The “Classification Label” column indicates the type of event (apnea or hypopnea), the “Signal Index” column specifies the index of the signal, and the “Segmented Signal Values” column contains the actual values of the segmented signal for that particular event.

By organizing the data in this manner, we can easily analyze and process the segmented signals for further analysis, such as feature extraction and classification tasks.

#### B. Feature extraction

Once the signals are denoised, decomposed, and segmented, feature extraction becomes crucial for signal classification. Several features are commonly extracted from input signals such as RR interval signals, ECG-derived respiration (EDR)

signals, heart rate variability (HRV), oxygen saturation signal (SpO2), blood gas or blood oxygen saturation (SaO2), and autocorrelation function (ACF), as reported by many researchers. The time-domain features considered in this study include: (i) the difference in the root mean square of RR peak amplitude, (ii) the time interval between consecutive HRV signals not exceeding 50 milliseconds, (iii) standard deviation of HRV signals, (iv) mean, variance, and kurtosis of ECG signals, (v) mean of ECG signals, and (vi) variance coefficient. These features provide valuable insights for characterizing and classifying sleep apnea signals. Most of the frequency domain features are extracted using Wavelet Transform (WT), as WT uses a multiscale basis, and it is advantageous over Fourier transforms [6].

After extracting the corresponding features for each signal, we constructed a second data frame to represent the features. The structure of the data frame is illustrated in “Fig. 2”.

No	Lefteye_mean	Lefteye_std	...	ECG_ch3_hrv_mean	ECG_ch3_hrv_std	Label
0	-0.003143	0.673188	...	0.014905	0.000	HYP-C
1	-0.04909	1.250325	...	0.008357	0.000	HYP-C
...	...	...	...	...	...	...
122	0.182651	0.809591	...	0.010316	0.000	APNEA-M
123	0.009362	1.118333	...	0.109668	0.000	APNEA-M

Fig. 2. Final data frame

In this data frame, each row represents a specific instance or segment of the signal, with features such as Lefteye-mean, Lefteye-std, ECG-ch3-hrv-mean, ECG-ch3-hrv-std, and so on. The “No” column indicates the index of the instance, while the “Label” column represents the corresponding classification label (HYP-C or APNEA-M).

By organizing the features in this structured data frame, it becomes easier to analyze and process the extracted features for classification tasks.

### C. Feature selection

For feature extraction section, we employed the “SelectKBest” method as a dimension reduction technique. This method is widely used in machine learning and feature selection tasks to identify the most informative features from a given dataset. The SelectKBest method evaluates the relevance of each feature by scoring them based on a specific statistical measure, such as chi-squared, mutual information, or ANOVA. It then selects the top-k features with the highest scores, where k is a user-defined parameter. This process helps to eliminate irrelevant or redundant features, improving the efficiency and effectiveness of our classification model. [7]. Specifically, we applied the SelectKBest method with a parameter k set to 50, indicating that we aimed to retain the top 50 most informative features for further analysis.

The “Fig. 3” illustrates the dataset before and after applying the SelectKBest method.

## V. CLASSIFIER

Random forests (RFs) are a combination of classification tree (CT). The classifier employs multiple decision trees, each

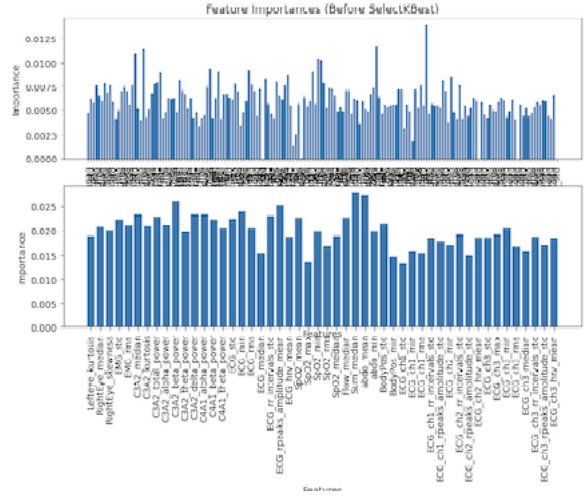


Fig. 3. SelectKBest

trained on random permutations of  $[\log_2 N]=3$  combinations of the N features. [5]. Each tree is trained on a different subset of the training set, and the final prediction output is averaged.

To ensure precise classification, we randomly divided the dataset, allocating 80 percent of the samples for training purposes and the remaining 20 percent for validation.

To ensure reproducibility and consistency, we set the random-state parameter to 42. This setting ensures that the same train-test split is generated consistently whenever the code is executed, enabling fair comparisons and reliable results.

The random forest (RF) classifier was trained to classify each signal into one of six states. After training, the classifier achieved an accuracy rate of approximately 76 percent, indicating its capability to correctly classify the signals with a favorable level of accuracy.

### A. Metrics for classification

The performance metrics that are used for the evaluation of the performance of the classifier are accuracy, sensitivity, and specificity. Accuracy refers to the closeness of the obtained result to that of the standard value. It has been calculated by measuring the true positives (TP), false positives (FP), true negatives (TN), and true positives (TP) using Equation “(1)”. Sensitivity and specificity are calculated using Equation “(2)” and Equation “(3)”, respectively. The precision can be calculated from Equation “(4)” [6].

$$Accuracy = 100 * (TP + TN) / (TP + TN + FP + FN), \quad (1)$$

$$Sensitivity = 100 * TP / (TP + FN), \quad (2)$$

$$Specificity = 100 * TN / (TN + FP), \quad (3)$$

$$Precision = TP / (TP + FN), \quad (4)$$

## VI. SHAP

Among other feature selection algorithms, cooperative Game Theory, in particular using the Shapley value, can be used to identify the most meaningful features that contribute the most to the decision process. Based on the general theory of the Shapley value, SHAP is a practical package in Python that is able to calculate the Shapley value for different ML models [5]. In this study, we utilized the Shapley value method on the output of a random forest classifier to determine the relative importance of each feature in contributing to the classification of apnea or hypopnea.

## VII. RESULTS

Next, we present an analysis of the outcomes derived from the Shapley values obtained for the trained model. The “Fig. 4” shows the significant features extracted from the data frame and their corresponding Shapley values, which indicate the magnitude of their impact on the model’s output.

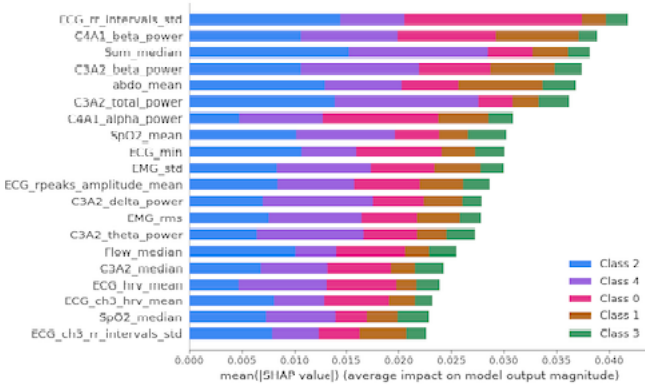


Fig. 4. Shap result

The graph yields the following results:

- The ECG R intervals standard deviation, C4A1 beta power, and sum median have the highest positive impact on the model’s output.
- The C3A2 beta power, abdominal mean, C3A2 total power, and C4A1 alpha power also contribute significantly to the model’s output.
- Other features such as SpO2 mean, ECG minimum, EMG standard deviation, ECG peaks amplitude mean, C3A2 delta power, and EMG root mean square (RMS) also have some level of impact, but to a lesser extent.
- The flow median, C3A2 median, ECG HRV mean, ECG ch3 HRV mean, SpO2 median, and ECG ch3 R intervals standard deviation have relatively lower impact on the model’s output.

The Shapley values provide insights into the importance of each feature in influencing the model’s predictions for different classes. Features with higher positive Shapley values contribute more towards predicting specific classes, while features with negative or lower Shapley values have less impact.

The heat map and correlation matrix are presented in “Fig. 5” and “Fig. 6”, respectively.

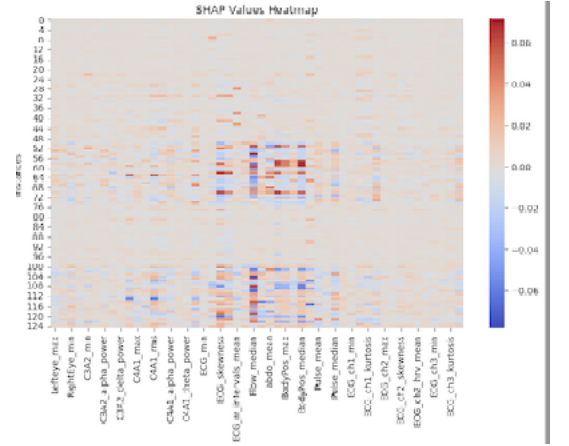


Fig. 5. Heat map

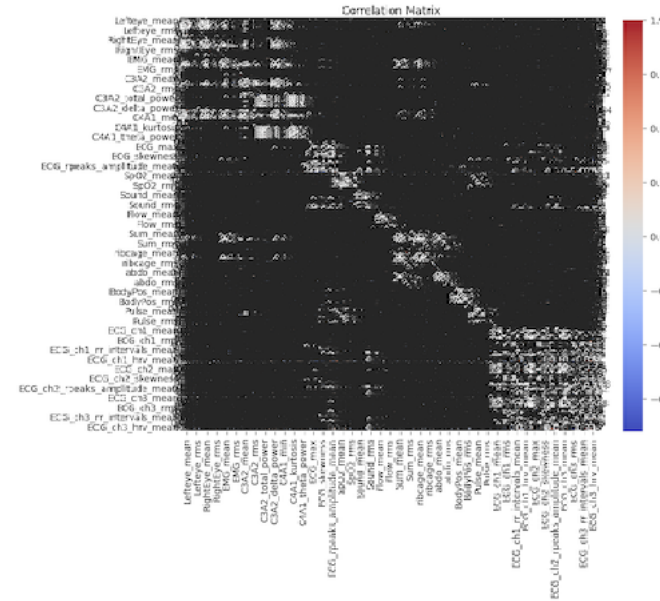


Fig. 6. Correlation matrix

## VIII. CONCLUSION

The Shapley value was employed to interpret the outcomes of a sleep disorder classifier with six categories. Our analysis revealed that certain signals pose a significant risk and carry the most influence in the classifier’s decision-making process. Notable examples include ECG R interval, C4A1 beta power, C3A2 beta power, abdominal mean and e.t.c., utilizing these identified features for classification purposes is anticipated to yield enhanced or comparable accuracy compared to alternative approaches.

## ACKNOWLEDGMENT

This work was supported by Dept. of Information Engineering (DEI) at the University of Padova,

## REFERENCES

## REFERENCES

- [1] Mencar C, Gallo C, Mantero M, et al. Application of machine learning to predict obstructive sleep apnea syndrome severity. *Health Informatics Journal*. 2020;26(1):298-317. doi:10.1177/1460458218824725
- [2] Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med* 2017; 13(3): 479–504.
- [3] A. T. Azar, H. I. Elshazly, A. E. Hassanien, and A. M. Elkorany, "A random forest classifier for lymph diseases," *Comp. Meth. Prog. Biomed.*, vol. 113, no. 2, pp. 465–473, 2014.
- [4] St. Vincent's University Hospital / University College Dublin Sleep Apnea Database. Available at: <https://physionet.org/content/ucddb/1.0.0/> (Accessed May 25, 2023).
- [5] D. Scapin, G. Cisotto, E. Gindullina and L. Badia, "Shapley Value as an Aid to Biomedical Machine Learning: a Heart Disease Dataset Analysis," 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), Taormina, Italy, 2022, pp. 933-939, doi: 10.1109/CCGrid54584.2022.00113.
- [6] JeyaJothi ES, Anitha J, Rani S, Tiwari B. A Comprehensive Review: Computational Models for Obstructive Sleep Apnea Detection in Biomedical Applications. *Biomed Res Int*. 2022 Feb 16;2022:7242667. doi: 10.1155/2022/7242667. PMID: 35224099; PMCID: PMC8866013.
- [7] Feature Selection with SelectKBest in Scikit-learn. Available at: [https://scikitlearn.org/stable/modules/generated/sklearn.feature\\_selection.SelectKBest.html](https://scikitlearn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html). (Accessed July 30, 2023).
- [8] Implemented python code Availabe at:  
[https://drive.google.com/file/d/1V8JRIYwPTKf\\_wUw4Gqq5bMiB1U8ExMcm/view?usp=drive\\_link](https://drive.google.com/file/d/1V8JRIYwPTKf_wUw4Gqq5bMiB1U8ExMcm/view?usp=drive_link)