

XGBoost Optimization using Single Lead ECG based on Multi-feature of ECG Morphologies for Automatic Sleep Disorder Classification

Iman Fahrudi

iman@polibatam.ac.id

Politeknik Negeri Batam

Muhamad Haikal Akmal

Politeknik Negeri Batam

Ridwan Ridwan

Politeknik Negeri Batam

Abdurahman Dwijotomo

Politeknik Negeri Batam

Research Article

Keywords: ecg, XGBoost, sleep disorder, multi feature, single lead ecg

Posted Date: August 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10453130/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

XGBoost Optimization using Single Lead ECG based on Multi-feature of ECG Morphologies for Automatic Sleep Disorder Classification

Iman Fahrudi^{1*}, Muhamad Haikal Akmal², Ridwan³, Abdurahman Dwijotomo⁴

^{1,2,3,4}Department of Electrical Engineering, Politeknik Negeri Batam, Batam, Indonesia, 29461

(*Corresponding author to provide e-mail: iman@polibatam.ac.id)

Abstract— sleep disorders arise from disturbances in the respiratory tract during sleep. These disturbances typically recur over time and occur abruptly. Their symptoms are commonly identified through polysomnography (PSG), the gold standard used by clinicians to diagnose sleep disorders. Among the signals acquired through PSG is the electrocardiogram (ECG), whose pattern is manually annotated by a clinician for further diagnosis. This procedure is time-consuming and highly complex, and it requires specialized expertise. To address this problem, a system that automatically classifies sleep disorders from single-lead ECG using a regularized XGBoost method is developed. The aim of this study is to develop a new sleep-disorder classification method based on XGBoost that exploits multiple features of ECG morphology to identify the signal patterns associated with sleep disorders. The proposed method evaluates the performance of a boosting ensemble in classifying apnea and non-apnea events. The parameter settings for XGBoost regularization were optimized to obtain the best performance, and the resulting accuracy was compared with that of other methods. The results show that the proposed method achieves an accuracy, specificity, and sensitivity of 96.1%, 100%, and 92.1%, respectively. The XGBoost classifier yields competitive results relative to other methods and is a powerful tool for automatic classification.

Index Terms— ecg, XGBoost, sleep disorder, multi feature, single lead ecg

I. INTRODUCTION

Sufficient and regular sleep is necessary to maintain a healthy and normal body condition. Good health helps maintain the body's hormonal balance and supports a positive state of mind. However, advances in information technology have reshaped modern lifestyles. Work can now be performed anywhere and at any time, leaving the body insufficient time to recover, increasing mental load, and imposing demanding schedules together with other external pressures. As a result, the body becomes fatigued and drowsy throughout the day, which in turn affects mental well-being [1]. The problems that subsequently arise include sleep disorders such as insomnia, sleep apnea, hypersomnia, REM sleep behavior disorder, circadian rhythm disorder, periodic limb movement disorder, shift-work sleep disorder, and narcolepsy [2],[3]. Such conditions can have serious consequences if they remain undiagnosed because of limited awareness of the need for medical examination [4]. In the long term, they may lead to a range of diseases, impair quality of life, and even result in death.

One common sleep disorder characterized by respiratory obstruction is obstructive sleep apnea (OSA). It is marked by partial or complete upper-airway obstruction during sleep, and these events can be both periodic and sudden [5]. Clinically, sleep apnea is categorized into three types: obstructive sleep apnea (OSA), central sleep apnea (CSA), and mixed sleep apnea (MSA). Its severity is determined from the apnea-hypopnea index (AHI), defined as the number of apnea and hypopnea events per hour. According to the standard criteria, an AHI below 5 is considered normal, an AHI of 5–14 indicates mild OSA, an AHI of 15–30 indicates moderate OSA, and an AHI above 30 indicates severe OSA [6]. The prevalence of OSA in the general population is estimated at 9%–24%, and as many as 90% of cases remain undiagnosed [7]. Patients with undiagnosed OSA therefore warrant close attention to reduce the risk of more serious complications.

In some cases, OSA sufferers are at increased risk of cardiovascular disorders [8], [9]. Further evidence indicates an association between respiratory distress and an increased likelihood of hypertension [10], as well as a relationship between snoring during sleep and stroke [11]. The rising number of OSA cases is attributable to many factors. For instance, OSA affects both men and women [12] and can occur at any age, from children and adolescents to the elderly. That is associated with obesity, smoking, and alcohol consumption [13]. Reported prevalence rates also vary because of differences in diagnostic tools, definitions, study designs, and subject characteristics [13].

One way to investigate signs of sleep disturbance is through signal analysis. Diagnosis is generally performed by monitoring airflow to detect the presence or absence of obstruction during sleep over a defined interval. When a more

detailed diagnosis is required, the findings are confirmed using additional signals recorded with a polysomnography (PSG) device. PSG can record a variety of signals, including those arising from brain activity [14], [15], cardiac activity [16], [17], [18] and muscle-contraction activity. Although PSG has become the standard tool for diagnosing sleep disorders, it also has several drawbacks. Among others are installation is complex because it requires 12 sensors and numerous cables attached to different parts of the body, interpreting the recordings demands specialized expertise, the diagnostic process is lengthy, the cost of a complete examination is high, access is limited by the small number of available PSG units, and patients find the procedure uncomfortable owing to the attached sensors and cables. These factors increase complexity and can degrade the quality of the recordings. As an alternative, a home sleep study (SSH) conducted under standardized clinical protocols is needed as a first-line service to obtain complete overnight recordings. An SSH device is typically simple, low-cost, and handheld. It uses only a single lead and minimal wiring, yet preserves the recording quality and clinical information required so that the ECG can be used by physicians for diagnosis. When necessary, any missing information can subsequently be obtained at the hospital to expedite decision-making and further medical action.

Feature extraction plays an important role in identifying the signal patterns of sleep disorders from relevant features. A detailed understanding of the dominant features helps reveal their correlation with the signal patterns of OSA patients. Selecting an appropriate method is therefore decisive for the success of the classification process. At the same time, the clinical information contained in the signal must be preserved so that physicians and other medical personnel can readily interpret the recordings and make subsequent clinical decisions. Over the past decades, several feature-extraction methods have been proposed. One introduced a new ECG-derived respiration (EDR) signal to reduce the distortion between EDR and RR intervals during apnea events on the basis of ECG morphology and respiratory activity [19]. Other studies used cardiac-rate points to derive linear and non-linear variables such as cepstral coefficients (CC), filterbanks (Fbank), and detrended fluctuation analysis [20]. Another study applied wavelet-packet analysis with Daubechies, Symlet, and biorthogonal wavelets to obtain the best basis function as a feature set [21]. To mitigate the shortcomings of standard ECG signal processing techniques, contemporary research underscores the significance of strong preprocessing systems, data augmentation methods, and thorough ablation evaluations. These approaches are designed to augment signal fidelity, diminish noise artefacts, and analyse the comparative influence of distinct features. Apart from preprocessing, the synthesis of spatial and temporal feature extraction has arisen as a favourable direction. Spatial representations embody morphological changes in ECG waveforms, while temporal modelling exploits sequential relationships between cardiac cycles. The amalgamation of these supportive perspectives delivers a more exhaustive characterization of physiological dynamics, consequently improving the capacity to discern apnea from non-apnea events[22].

Classifying sleep disorders is a complex task that requires an appropriate approach to obtain accurate, clinically defensible results on which experienced physicians and other medical personnel can rely. One study achieved an average accuracy of 90.27% in classifying sleep apnea into three types, i.e., obstructive, central, and mixed, using wavelet-coefficient values selected through a feature-selection process based on a support vector machine [23]. A combination of EDR and RR-interval features extracted from one-minute ECG recordings in the PhysioNet database achieved a classification accuracy of 90.7% [19]. Several SVM variants were evaluated, and the radial basis function (RBF) kernel achieved the highest accuracy, 84.74%, on 60-second ECG recordings from the PhysioNet database [18]. Furthermore, a quadratic discriminant analysis (QDA) method using linear and non-linear features derived from the cardiac cycle attained an accuracy of 84.76% on five-minute ECG recordings from the Apnea-ECG PhysioNet database [20]. The use of deep learning to automatically extract features, combined with a decision-fusion classifier, yielded an accuracy of 84.7% on the Apnea-ECG PhysioNet database with a 60-second segment duration [24].

The growing body of research on automated systems for identifying and classifying sleep disorders is motivated mainly by two objectives. The first is to reduce human intervention in the assessment, because human interpretation is inherently limited and variable, as reflected in intra-observer and inter-observer validation tests [25],[26]. A clear and definitive standard is therefore needed to diagnose sleep disorders from ECG patterns, which represent cardiac function during sleep. Detection algorithms can be standardized provided that they achieve high sensitivity, specificity, and accuracy, so that they can support medical decision-making and complement the clinical information obtained from PSG and home sleep studies. The second objective addresses a limitation of previous studies, which focused only on cardiac rhythm in the form of RR-interval variability without analyzing the periodic pattern and morphology of the ECG signal. Since both are needed to obtain complete clinical information, the accuracy of the decision-making algorithm must be improved while maintaining computational efficiency and high precision.

This paper proposes a new framework that classifies patients as either non-apnea (healthy) or apnea. The framework has four goals. First, it assists clinical experts in diagnosing sleep disorders using a prediction algorithm for suspected patients in the sleep clinic. Second, it predicts the presence or absence of sleep-disorder events using a multi-feature analysis based on ECG morphology and heart rate variability. Third, it identifies periodic patterns in the ECG signal using only a single lead. Fourth, it develops an expert system that classifies sleep disorders with improved performance, supporting clinical decision-making and improving patient health.

II. SLEEP DISORDER CLASSIFICATION FRAMEWORK

This section describes the prediction system, which takes a set of features and analyzes them to reach a final decision. The proposed framework for sleep-disorder classification based on multi-feature analysis (MFA) using XGBoost is shown

in Fig. 1. The scheme consists of three stages. The first stage is preprocessing, which produces ECG recordings free of noise. The second stage is feature extraction, in which the ECG recordings are searched for features relevant to the signal patterns of sleep disorders. The third stage is classification into two classes: the non-apnea class (healthy) and the apnea class (OSA). The classifier used is XGBoost, a member of the ensemble boosting family.

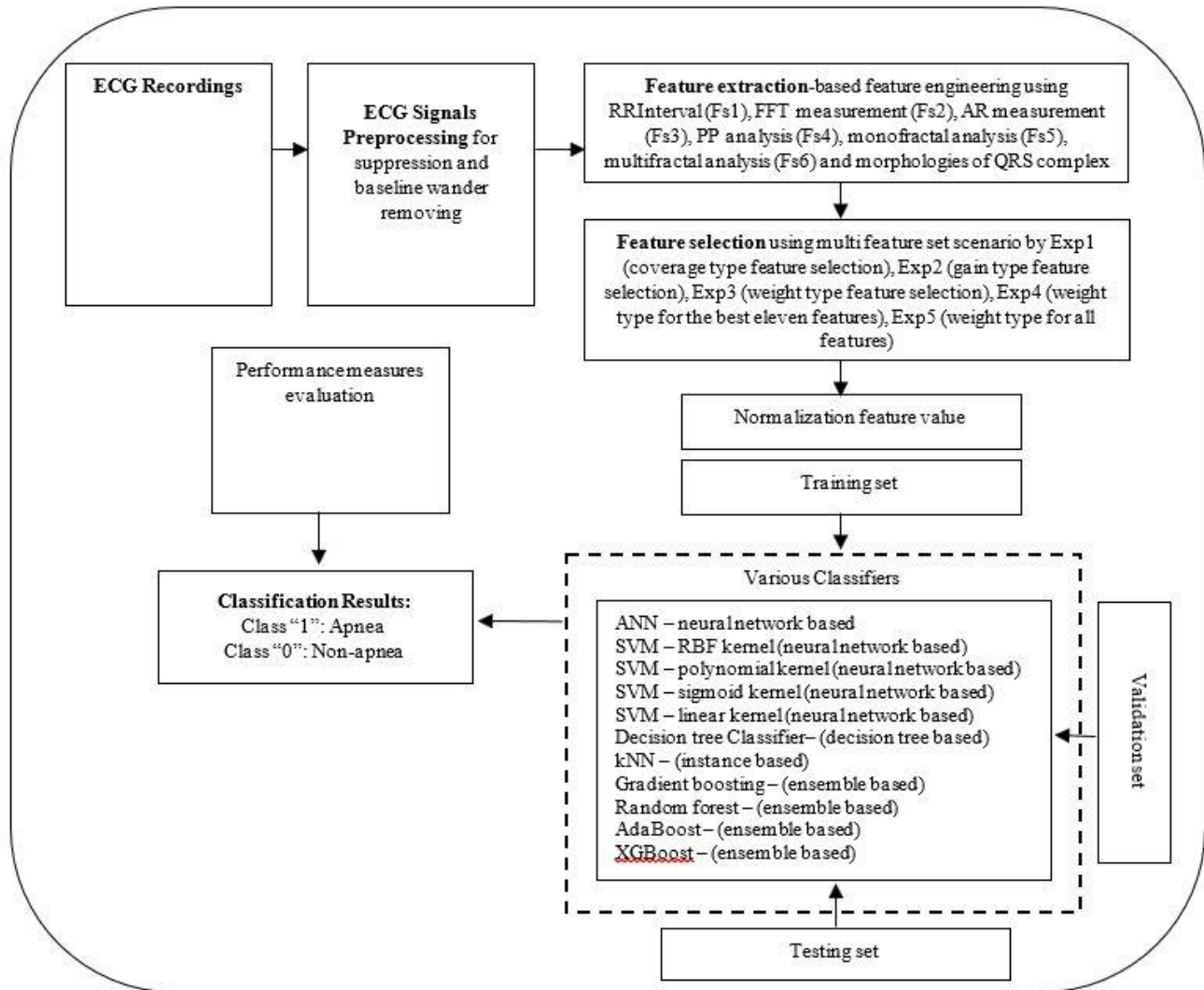


Fig. 1. The framework of classification process to classify sleep disorder

2.1 ECG-Apnea Dataset

This study uses a dataset sourced from the physionet ECG-apnea database [27]. Each recording was labeled minute by minute as apnea (A) or non-apnea (N) by a medical expert. Whereas previous studies typically used test recordings of 30 seconds to 1 minute, this study used longer recordings of approximately 16 minutes each. Longer recordings are expected to contain richer clinical information, enabling the system to translate the resulting feature values more accurately and to distinguish the periodic patterns of the ECG signal for both non-apnea and apnea cases, thereby improving classification accuracy.

The test set comprised 343 ECG recordings with an equal number of non-apnea and apnea recordings, each characterized by 26 features. The feature values were obtained from the RR interval, the fast Fourier transform, autoregressive spectral analysis, the Poincaré plot, monofractal detrended fluctuation analysis [28], multi-fractal analysis [29], [30], multifractal analysis, and the QRS-complex morphology. Fig. 2(a) shows an example of a raw ECG signal for an apnea event, and Fig. 2(b) shows the corresponding denoised ECG signal.

2.2 Preprocessing Single Lead ECG

The proposed framework achieves high accuracy only if each stage produces high-quality records that provide complete clinical information without distortion. One prerequisite is preprocessing the ECG signal before feature extraction. Although the ECG recordings obtained from PhysioNet are not expected to contain severe noise introduced during acquisition, preprocessing is still necessary to remove unwanted components and baseline-wander noise.

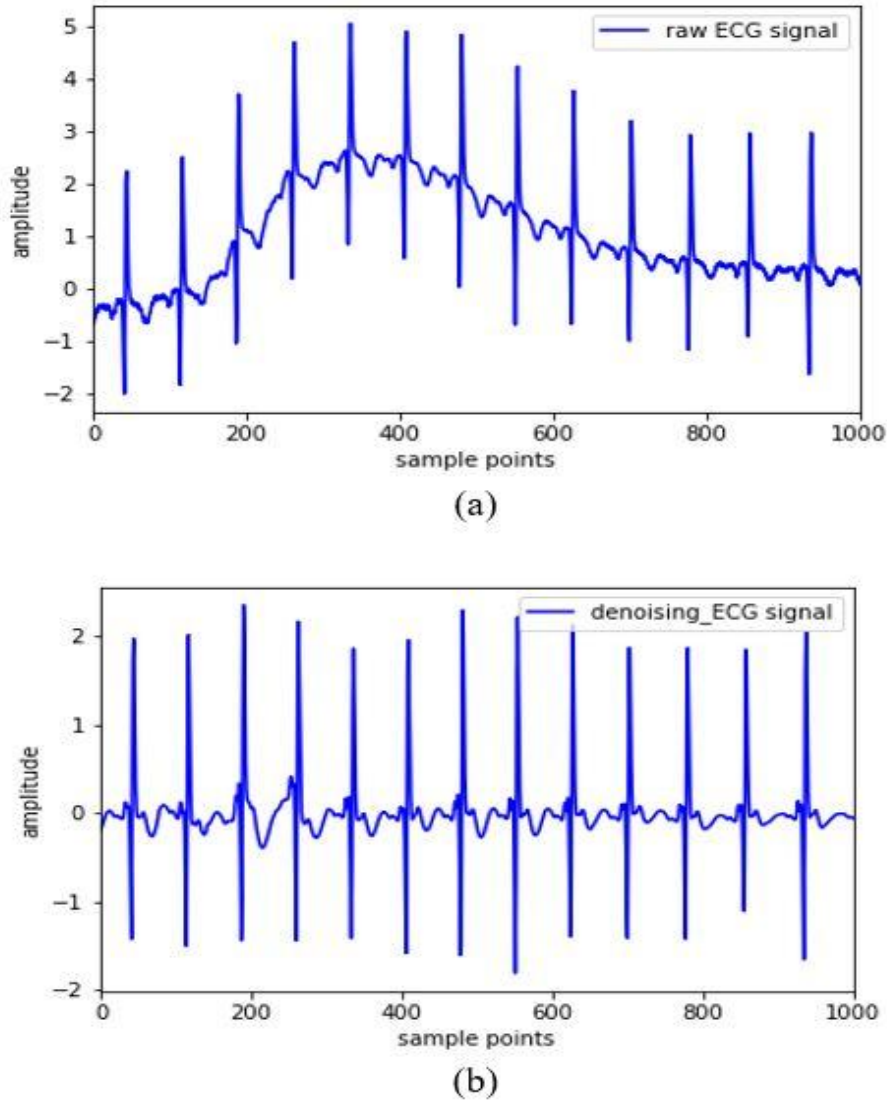


Fig. 2. Represent of single lead ECG duration ± 16 minutes; (a) An example of ECG signal for apnea event; (b) An example of denoising of ECG signal

Every recording acquired with PSG or other medical instruments is susceptible to low- and high-frequency noise from various sources, including powerline interference, lead-contact interference, wiring interference, muscle contraction and patient respiration, among others. Mathematically, the ECG signal $y[n]$ is defined with (1). The variable $s[n]$ is the representation of the ECG signal and $w[n]$ is the value obtained from the overall noise that interferes with the ECG signal. While α is the signal attenuation parameter.

$$y[n] = \alpha s[n] + w[n] \quad (1)$$

The QRS complex retains its shape even in the presence of substantial noise. Baseline-wander noise, however, strongly affects R-wave detection and hence RR-interval extraction, as it introduces errors. The first step is therefore to attenuate the low-frequency noise and restore the baseline to the zero (isoline) level. These steps are applied to all training and testing ECG records, yielding filtered signals ready for the subsequent QRS-detection phase.

As the most important component of ECG morphology, the QRS complex carries substantial clinical information because it reflects the electrical activity of the heart. This activity reveals when events occur, how they change, what they mean clinically, and how they relate to patients with sleep disorders. In ECG research, the detection method is always central to efforts to develop algorithms with high accuracy. The QRS complex has a frequency between 5 and 25 Hz, whereas the other waves lie below this range [31]. A widely used benchmark in ECG research, and the algorithm adopted in this study for QRS detection, achieves a high accuracy of 99.3% [32]. Another study used a deep-learning model based on a ConvNet architecture to detect the QRS complex with 99.62% accuracy [33]. QRS wave frequency range between 5 to 25 Hz [31]. The other waves have a frequency range below the frequency range of the QRS wave.

2.3 Feature Extraction

The most critical step is feature extraction, which derives feature values from the morphology and periodic pattern of the ECG signal. A successful extraction process yields accurate predictions and fast computation. The features are organized into seven sets. Feature set 1 comprises statistical measures of the RR interval, the inter-beat interval obtained throughout the overnight ECG recording. Feature set 2 is obtained with the fast Fourier transform (FFT) to estimate the distribution of power across frequency components. Feature set 3 represents power-spectral measurements obtained with the autoregressive (AR) method to estimate the distribution of power across the frequency spectrum. Feature set 4 is derived from a non-linear Poincaré plot (PP) analysis, which provides quantitative descriptors of short- and long-term recording patterns. Feature set 5 is obtained by an entropy-based technique, in which features are measured from the slope estimated by minimizing the local-trend error over small and large sample sizes using mono-fractal detrended fluctuation analysis (DFA). Feature set 6 is obtained from another entropy-based approach, multifractal detrended fluctuation analysis (MFDFA), which measures the width of the fractal structure from the minimum and maximum segment sample sizes. Finally, feature set 7 combines the varying dynamics of the inter-beat intervals with the QRS-morphology characteristics of the ECG signal, completing the set of features that represent apnea and non-apnea signal patterns. Together, these features enable the machine-learning model to distinguish the two classes more easily and thus improve performance. The feature sets are described below:

1) Feature set 1 (fs1);

Heart-rate variability (HRV) analysis based on the RR interval. During sleep, some parts of the body rest while others, such as the heart, continue to work. The electrical activity of the heart is recorded as an electrocardiogram (ECG). Accordingly, the measurement device must capture clinical information as a time series in fine detail. These variations are described by several parameters computed from the time intervals between R peaks. This feature set comprises RR_mean , RR_std , the root mean square of successive RR-interval differences (RMSSD), and the triangular interpolation of the NN interval (TINN).

2) Feature set 2 (fs2);

In addition to time-domain measurements, ECG recordings can be assessed in the frequency domain. A powerful tool for this purpose is the fast Fourier transform (FFT). Because ECG signals differ in frequency, amplitude, and phase, the Fourier series is well suited to decomposing them into frequency bands. This technique enables computation of the absolute power in the very-low-frequency (VLF), low-frequency (LF), and high-frequency (HF) bands from the spectral distribution relative to the peak frequency of the power spectral density (PSD). These parameters were computed with an FFT configured for 2 Hz interpolation and a frequency bin of 1024, using a Hanning window of length 1024.

3) Feature set 3 (fs3);

Another frequency-domain approach used to assess the ECG signal is the autoregressive (AR) spectrum. This technique produces a PSD estimate from which the absolute power in the VLF, LF, and HF bands is obtained. These three parameters are computed in three steps: resampling the RR-interval signal, estimating the PSD from the resampled signal, and computing the frequency-domain parameters from the resulting PSD. To estimate the peak frequency of the PSD, the AR model used a forward order of 16, an interpolation rate of 2 Hz, and a frequency bin of 1024.

4) Feature set 4 (fs4);

This feature set is obtained through the non-linear Poincaré plot (PP) technique, which evaluates the distribution of points as an ellipse defined by two intersecting axes for the RR time series RR1, RR2, ..., RRn, RRn+1. The PP yields two descriptors, SD1 and SD2. SD2 is the standard deviation of the point distribution between RRn and RRn+1 projected onto the identity line ($y = x$), whereas SD1 is the standard deviation along the line perpendicular to the identity line.

$$d_{1,i} = \frac{1}{\sqrt{2}} |X(i) - X(i + 1)|, i = 1, 2, \dots, N - 1 \quad (2)$$

$$SD1 = \sqrt{\text{var}(d1, i)} \quad (3)$$

$$d_{2i} = \frac{1}{\sqrt{2}} |X(i) - X(i + 1) - 2\bar{X}| \quad (4)$$

$$SD2 = \sqrt{\text{var}(d2, i)} \quad (5)$$

5) Feature set 5(fs5);

The ECG signal varies continuously, and these variations can be analyzed to characterize the underlying signal dynamics. One entropy-based method evaluates self-similarity through local detrending within defined segments. Detrended fluctuation analysis (DFA) is well suited to estimating the slope, commonly known as the scaling exponent, obtained from local detrending and described by the parameters alpha1, residue1, alpha2, and residue2. The mono-fractal analysis algorithm used in this work is implemented as follows to estimate the signal slope for apnea and non-apnea segments.

Step 1: Determine RR(i), the RR(i) is RR interval of the i th.

Step 2: Calculate RR (ave), it is mean of RR interval.

Step 3: Calculate the segmented profile of signal $y(i)$ of length N as described in (6).

$$y(i) = \sum_{k=1}^i RR(k) - RR(ave), i = 1, \dots, N \quad (6)$$

Step 4: Segmented the $y(i)$ into $N_t = \text{int}(N/t)$ segments. To complete the segmented process by calculate into $2N_t$.

Step 5: Subtract the local trend for each segment. Calculate the local trend by a least square fit, and determine of variance $v, v=1, \dots, N_t$. The parameter $p_v(i)$ is fitting process in the segment v .

$$Ft^2(v) \equiv \frac{1}{t} \sum_{i=1}^t [y(v-1)t + i - p_v(i)]^2 \quad (7)$$

Step 6: Root-mean-square is calculated by:

$$F(t) = \left[\frac{1}{2N_t} \sum_{v=1}^{2N_t} Ft^2(v) \right]^{1/2} \quad (8)$$

Step 7: Repeat procedure 2 and 6 to obtain the $\log F(n)$ and $\log n$ correlations

Fig 3 shows the relationship of $\log F(n)$ and $\log(n)$. The scaling exponent is an indicator of correlations. Values of $\alpha = 0 - 0.5$ indicate anti-correlations for short range in the time series, value of $\alpha = 0.5 - 1$ indicate correlated for long range structure.

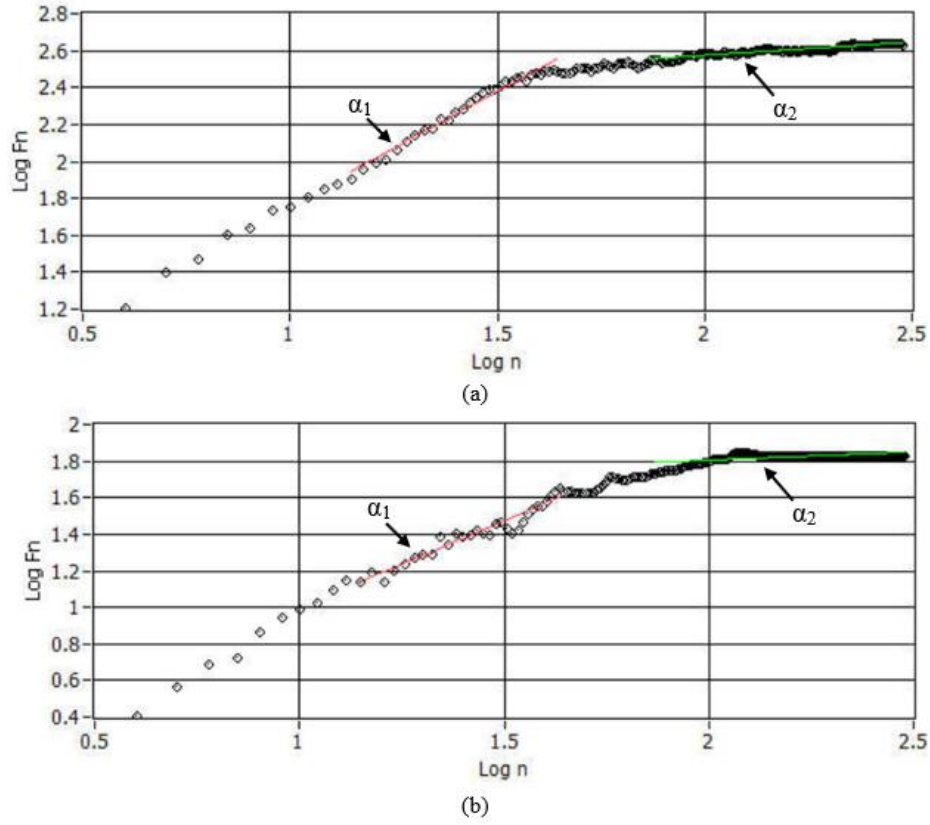


Fig. 3. Represent of plot the relationship of log F(n) and log(n); (a)-(b) An example of the scaling exponent(α) for apnea events

6) Feature set 6(fs6);

Beyond the slope obtained from mono-fractal analysis, the signal pattern can also be characterized by its scaling behavior, which captures the detail of each scaling change. The procedure consists of determining the signal profile, partitioning it into non-overlapping segments, locally detrending each segment, computing the fluctuation function, and determining the scaling behavior, namely the singularity exponent ($h(q)$) with its maximum and minimum and the singularity dimension ($D(q)$) with its corresponding maximum and minimum. The multifractal analysis algorithm used in this paper is implemented as follows to compute the fractal dimension ($D(q)$), as described in (9), (10), and (11). The algorithm demonstrates the discriminative power of multifractal analysis for apnea and non-apnea signals.

$$F_q(t) \sim t^{h(q)} \quad (9)$$

$$t(q) = qh(q) - 1 \quad (10)$$

Whereas the $D(q)$ is obtained by Eq. (11).

$$D(q) = \frac{t(q)}{q-1} = \frac{qh(q)-1}{q-1} \quad (11)$$

Consequently, every change in the parameter q will affect the fractal dimension ($D(q)$). An example of $D(q)$ and $h(q)$ display the multifractal spectrum in Fig 4.

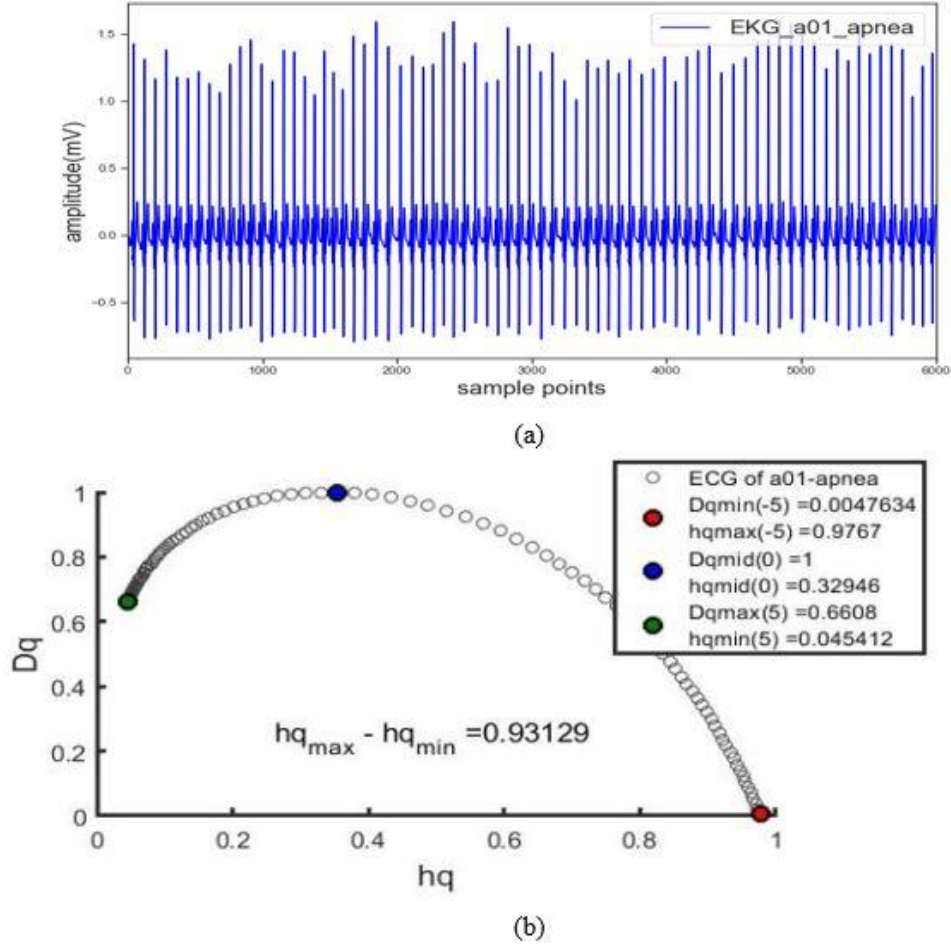


Fig. 4. Represent of plot multifractal spectrum (a) An example of ECG signal for apnea event; (b) An example of multifractal spectrum

7) Feature set 7 (fs7);

The heart contracts as a result of depolarization of the cardiac muscle. This contraction is reflected in changes in the shape of the recorded ECG signal. ECG morphology is characterized by the P wave, the QRS complex, and the T wave. Fig. 5 shows the ECG waveform and the characteristics of the QRS-complex feature set. These waveform changes manifest as differences in shape, width, and height and are clinically related to depolarization of the atria and ventricles. The set of features are QRS amplitude mean (QRSam), QRS amplitude standard deviation (QRSas), QRS width mean (QRSwm) and QRS width standard deviation (QRSws).

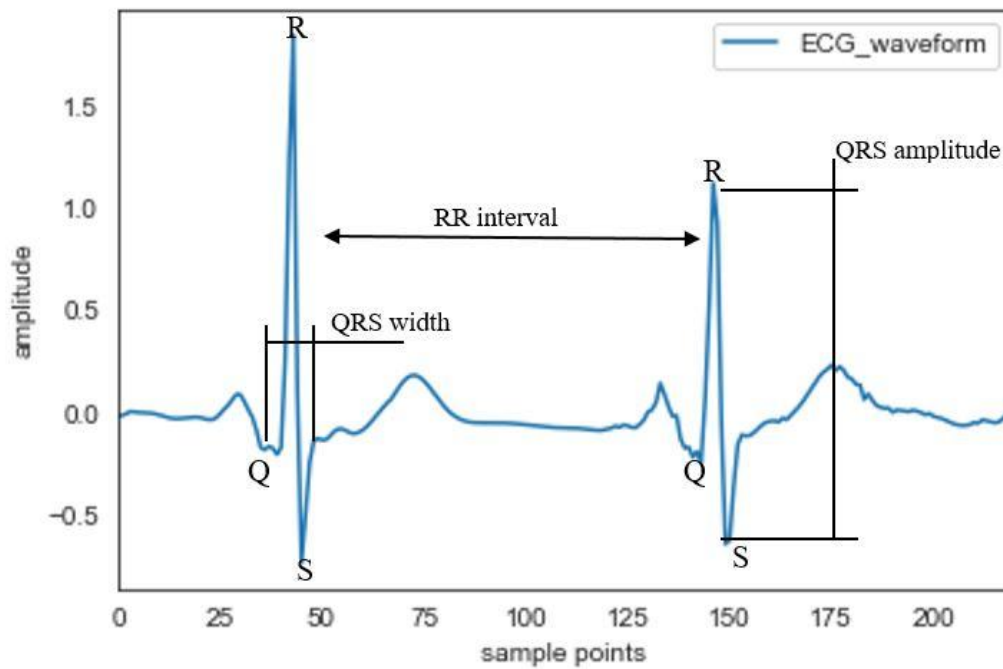


Fig. 5. ECG waveform and various of interval

2.4 Calculation of Feature Importance for Fitting to XGBoost Classification Model

This step is used to remove uninformative features. Feature-importance estimation also serves as a mechanism by which the model ranks and selects the most useful features, aids interpretation of the dataset, improves model structure, and reduces dimensionality, thereby increasing the model's effectiveness. Technically, feature importance is computed by assigning a score to each feature, and the highest-scoring features are used as model inputs. These scores are derived from the tree structure produced by the gradient-boosted decision-tree scheme, in which each tree generates nodes, and each node corresponds to a single feature. Two broad approaches are used to score features: a regression approach, which scores features from numerical data for prediction, and a classification approach, which scores features from the labeled dataset. In XGBoost, feature importance can be computed in several ways based on gradient-boosted decision trees, including weight, gain, and coverage. The weight is the number of times a feature is used to split the data across all trees, the gain is the average gain of the splits in which the feature is used, and the coverage is the average coverage of the splits in which the feature is used.

2.5 Extreme Powerful XGBOOST Machine Learning for Classification Sleep Disorder

The XGBoost algorithm is based on gradient boosting machine learning and stands for Extreme Gradient Boosting [34], [35]. The XGBoost algorithm belongs to the boosting algorithm family and operates as supervised machine learning to solving the prediction and classification problems. One of the differences between gradient boosting and XGBoost is that it uses a regularization technique to reduce overfitting. Another advantage of XGBoost over traditional gradient boosting is its high execution speed and support for parallel and distributed processing. Boosting is a decision-tree scheme that operates sequentially, correcting the errors of previous models so that a strong learner is ultimately built from a sequence of weak learners.

To this end, a multi-feature training set is used to classify non-apnea and apnea events. The input features, target variable, objective function, and number of iterations constitute the basic parameter settings for XGBoost. In this work, the model is built in six phases to classify ECG-apnea recordings into two categories, apnea and non-apnea, as follows:

- Phase 1: import classes and function.
- Phase 2: load dataset and split dataset into training set and testing set.
- Phase 3: Fitting the XGBoost model on training set.
- Phase 4: get the importance feature
- Phase 5: Evaluation the XGBoost model on testing set
- Phase 6: Evaluation prediction performance the XGBoost model.

2.6 Training and Evaluation Model

The dataset is then split into a training set and a testing set. The training set serves as input for training the XGBoost model, while the testing set is used to evaluate the model so that it is able to predict based on the modelling obtained previously. In this process, to determine the reliability of the proposed model, the dataset is divided into 288x25 matrix with 7200 entries for the training set loaded, while the testing set is 96x25 matrix with 2400 entries loaded. The performance of XGBoost classifier is being evaluated in four evaluation parameters as follows:

$$\text{Accuracy(Ac)} = \left(\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \right) \times 100 \quad (12)$$

$$\text{Sensitivity(Se)} = \left(\frac{\text{TP}}{\text{TP} + \text{FN}} \right) \times 100 \quad (13)$$

$$\text{Specificity(Sp)} = \left(\frac{\text{TN}}{\text{TN} + \text{FP}} \right) \times 100 \quad (14)$$

$$\text{Precision(P)} = \left(\frac{\text{TP}}{\text{TP} + \text{FP}} \right) \times 100 \quad (15)$$

In these equations, a true positive (TP) is an apnea event correctly classified as apnea, and a true negative (TN) is a non-apnea event correctly classified as non-apnea. Likewise, a false positive (FP) is a non-apnea event misclassified as apnea, and a false negative (FN) is an apnea event misclassified as non-apnea. The numbers of TP, TN, FP, and FN cases are obtained from the confusion matrix for apnea–non-apnea classification. Table 4 lists the scenarios and parameter settings evaluated. Each model is configured and optimized according to the characteristics of the dataset and the classification objective in order to obtain the most accurate model.

2.7 Parameter tuning for Optimization

Implementing XGBoost with gradient boosting requires tuning several parameters to obtain optimal results. These parameters fall into three groups: general parameters governing the overall modelling process, booster-specific parameters, and learning-task parameters. The general parameters include the booster type, which controls how the model operates; the silent mode, which aids model inspection; and nthread, which enables parallel processing. The booster-specific parameters include eta (the learning rate) and min_child_weight and max_depth, which help prevent overfitting. The learning-task parameters include the loss function and the performance metric. Parameter optimization aims to minimize the error between predicted and actual values. The three parameter groups are described in Table 4, and the general procedure for determining the hyperparameters of the XGBoost model is as follows:

- 1) Booster parameters setting: adjust the number of max depths in XGBoost model.
 - 2) Learning parameters: adjust the number of learning rate in XGBoost model.
 - 3) Regularization parameters: adjust the number of max depths, learning rate and other general setting in XGBoost model.
- The max_depth and learning rate were tuned individually and jointly to determine the relationship between tree depth and learning rate during training, evaluation, and testing, so as to obtain optimal results. The remaining parameters were left at their default XGBoost values.

III. RESULTS

This section presents the results of the feature-extraction process to assess the effectiveness of the feature set as input and to examine how the XGBoost classifier achieves high accuracy. The feature sets were compared to identify those most relevant to the classification task and to evaluate the performance of the proposed framework for sleep-disorder classification. Two processes are central: feature extraction, which assesses each ECG record to derive and discriminate feature values for apnea and non-apnea; and experimentation with various multi-feature scenarios to obtain the best classification accuracy.

The XGBoost framework operates in parallel using a tree-boosting mechanism to solve data-science problems accurately and efficiently; this study uses gradient-boosting regularization. By default, XGBoost provides a baseline set of parameters, but these are not necessarily suited to the amount of data, the feature set, or the number of classes. Experimentation and parameter optimization are therefore required to control the algorithm's behavior during training and to produce a high-performing model. Improved performance, in turn, requires statistically significant features. Table 2 reports the distribution of feature values (mean and standard deviation, SD) and the significance level for all features. All features in feature sets 1 and 4 have lower means and standard deviations in non-apnea than in apnea recordings, and the differences are highly statistically significant. These feature sets therefore contribute substantially to classification performance. Overall, 76.92% of the features are highly significant, and a further 11.54%, i.e., the hqmin, Dqmin, and hqmaxhqmin features in feature set 6 are significant. Thus, 88.46% of the features are statistically significant and are expected to improve classification performance. The remaining 11.54% high-frequency power (HF), residue1, and hqmax are not significant.

TABLE 1.
DESCRIPTION OF FEATURE SET

Feature set (FS)	Clinical Feature Characteristics	Descriptions
Feature set 1(fs1); The heartrate variabilities of RR interval calculation	f0: RR_mean f1: RR_std f2: RMSSD f3: TINN	Mean of RR intervals Standard deviation of RR interval Representative of beat-by-beat variance in the root mean square of successive differences between normal heartbeats RR interval histogram based the baseline width of the RR interval
Feature set 2(fs2); fast Fourier transform(fft) measurement	f4: Power of very low frequency (VLF) f5: Power of low frequency (LF) f6: Power of high frequency (HF)	The power absolute signal density of ECG signal in low to high frequency spectrum using FFT method
Feature set 3(fs3); autoregressive (AR) spectrum measurement	f7: Power of very low frequency (VLF) f8: Power of low frequency (LF) f9: Power of high frequency (HF)	The power absolute signal density of ECG signal in low to high frequency spectrum using AR method
Feature set 4(fs4); the Poincare plot analysis	f10: SD1 f11: SD2	Point distribution along the line by fitting an ellipse form
Feature set 5(fs5); The mono-fractal analysis based self-similarity of ECG Signal	f12: Alpha1 f13: Residue1 f14: Alpha2 f15: Residue2	Representative of short term and long-term fluctuations non-overlapping ECG signal
Feature set 6(fs6); Analysis of behaviour of ECG signal based multi-fractal calculation	f16: hqmin f17: hqmax f18: hqmid f19: Dqmin f20: Dqmax f21: hqmaxhqmin (Multifractal spectrum width)	Reflect of ECG behaviour in multi-fractal spectrum to calculate scaling behaviour(hq) with maximum and minimum variances, and singularity dimension (Dq) with corresponding maximum and minimum
Feature set 7(fs7); morphologies characteristic of QRS complex	f22: QRS amplitude mean (QRSam) f23: QRS amplitude standard deviation (QRSas) f24: QRS width mean (QRSwm) f25: QRS width standard deviation (QRSws)	Calculation QRS amplitude and QRS width based on fudicial point of ECG waveform

3.1 The contribution of selection feature

After feature extraction, each feature yields a value that is meaningful only insofar as it contributes to the success of the classifier. As noted, 26 features are obtained through various approaches suited to the characteristics of ECG signals. To identify the dominant features, a score is computed for each feature so that the features can be ranked accordingly. To quantify the contribution of a feature to the construction of the decision tree, model, XGBoost has a feature importance scheme. This scheme ranks each feature by how frequently it is used to build the decision tree. Fig. 6 shows the XGBoost feature-importance scores. For the gain type, the features in descending order of importance are TINN, RR std, LF power_ar, alpha2, hqmin, QRS width std, RR mean, QRS width mean, QRS amplitude mean, QRS amplitude std, Dqmin, alpha1, hqmid, residue1, hqmax, HF power_fft, hqmaxhqmin, VLF power_ar, RMSSD, SD2, LF power_fft, residue2, VLF power_fft, hqmax, HF power_ar, and SD1. For the weight type, the order is TINN, QRS width std, alpha1, Dqmax, QRS width mean, residue2, QRS amplitude std, Dqmin, RR mean, alpha2, hqmid, hqmax, hqmaxhqmin, SD2, VLF power_fft, residue1, RMSSD, VLF power_ar, HF power_ar, LF power_ar, QRS amplitude mean, LF power_fft, HF power_fft, hqmin, RR std, SD1. Finally, for the coverage metric, the feature order is TINN, QRS amplitude std, alpha1, LF power_ar, QRS amplitude mean, alpha2, QRS width std, QRS width mean, VLF power_ar, hqmaxhqmin, residue1, hqmid, HF power_ar, Dqmin, SD2, RR mean,

RR std, residue2, Dqmax, hqmin, LF power_fft, RMSSD, hqmax, HF power_fft, VLF power_fft, SD1. Therefore, to assess the effect of feature selection on the model configuration obtained with the recommended parameter settings, the test scenario is carried out according to the information obtained from Table 3, Table 4 and Table 5.

TABLE 3.
SUMMARIZING THE CALCULATED BEST FIVE FEATURE IMPORTANCE SCORES

Type of feature important algorithm	Feature quantifying
Weight type	TINN, QRS width std, alpha1, Dq max, QRS width mean
gain type	TINN, RR std, LF power, Alpha2, hq min
coverage type	TINN, QRS amplitude std, LF power, alpha1, QRS amplitude mean

Table 3 shows the recapitulation of feature importance using decision tree gradient boosting scheme on XGBoost Classifier. The test uses three types of feature selection owned by XGBoost based on score calculation. This feature selection is done to get features that have very dominant characteristics and are useful as input to a model. Each type selected 5 features that have the best score from 26 available features. Based on testing, different selected features are obtained in each test on type weight, gain and coverage. Testing on type weight resulted in five features with dominant scores, namely TINN, alpha1, QRS width

TABLE 2.
THE FEATURE VALUE OF FEATURE SET 1, 2, 3, 4, 5, 6 AND 7 WERE GIVEN FOR APNEA AND NONAPNEA (MEAN±STANDARD DEVIATION) AND THE SIGNIFICANCE LEVEL (P-VALUE)

ECG Feature set 1 (fs1)			
ECG Features	Feature values (Apnea)	Feature values (Non-apnea)	p-value
f0	918.74(89.46)	859.72(115.88)	<0.001
f1	105.34(48.38)	67.04(41.81)	<0.001
f2	104.36(69.44)	79.49(62.79)	<0.001
f3	272.379(114.72)	148.74(61.79)	<0.001
ECG Feature set 2 (fs2)			
ECG Features	Feature values (Apnea)	Feature values (Non-apnea)	p-value
f4	73.77(49.21)	56.24(44.40)	<0.001
f5	143.55(59.89)	92.66(45.92)	<0.001
f6	4579.17(5629.71)	1195.05(2067.15)	<0.001
ECG Feature set 3 (fs3)			
ECG Features	Feature values (Apnea)	Feature values (Non-apnea)	p-value
f7	4786.68(5430.72)	2401.46(3563.10)	<0.001
f8	3346.35(4690.16)	2193.58(3758.31)	<0.001
f9	7582.97(10207.41)	2208.85(4367.84)	≥0.05
ECG Feature set 4 (fs4)			
ECG Features	Feature values (Apnea)	Feature values (Non-apnea)	p-value
f10	7362.50(7129.27)	4423.91(6757.64)	<0.001
f11	4989.73(6751.67)	3940.87(6045.48)	<0.001
ECG Feature set 5 (fs5)			
ECG Features	Feature values (Apnea)	Feature values (Non-apnea)	p-value
f12	1.2216(0.3621)	0.9277(0.2666)	<0.001
f13	0.00059(0.00075)	0.00055(0.00070)	≥0.05
f14	0.4686(0.2471)	0.7037(0.2445)	<0.001
f15	0.00074(0.00084)	0.0012(0.0012)	<0.001
ECG Feature set 6 (fs6)			
ECG Features	Feature values (Apnea)	Feature values (Non-apnea)	p-value
f16	769.26(7622.28)	170.84(2353.15)	0.01 – 0.05
f17	1.4478(0.2358)	1.3834(0.2371)	≥0.05
f18	0.4685(0.07059)	0.4427(0.0951)	0.001 – 0.01
f19	0.6020(0.2729)	0.6858(0.3161)	0.01 – 0.05
f20	0.0856(0.1176)	0.1006(0.1284)	0.001 – 0.01
f21	1.3743(0.2766)	1.3070(0.2619)	0.01 – 0.05
ECG Feature set 7 (fs7)			
ECG Features	Feature values (Apnea)	Feature values (Non-apnea)	p-value
f22	1.8171(1.0242)	2.1657(1.0264)	<0.001
f23	0.6134(0.3819)	0.7713(0.4445)	<0.001
f24	109.21(28.28)	98.28(20.98)	<0.001
f25	67.10(21.74)	48.17(21.58)	<0.001

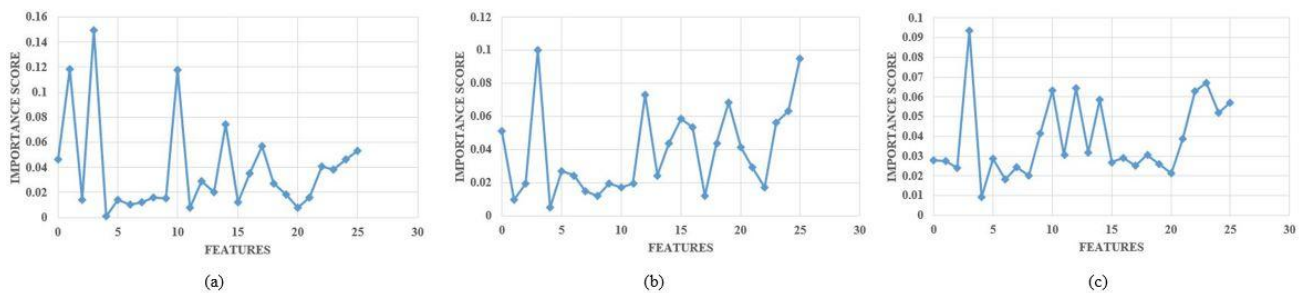


Fig. 6. The distribution of feature importance score; (a). The variation of score for gain type, (b). The variation of score for weight type, and (c). The variation of score for coverage type.

TABLE 3.
SUMMARIZING THE CALCULATED BEST FIVE FEATURE IMPORTANCE SCORES

Type of feature important algorithm	Feature quantifying
Weight type	TINN, QRS width std, alpha1, Dq max, QRS width mean
gain type	TINN, RR std, LF power, Alpha2, hq min

Type of feature important algorithm	Feature quantifying
coverage type	TINN, QRS amplitude std, LF power, alpha1, QRS amplitude mean

Table 3 shows the recapitulation of feature importance using decision tree gradient boosting scheme on XGBoost Classifier. The test uses three types of feature selection owned by XGBoost based on score calculation. This feature selection is done to get features that have very dominant characteristics and are useful as input to a model. Each type selected 5 features that have the best score from 26 available features. Based on testing, different selected features are obtained in each test on type weight, gain and coverage. Testing on type weight resulted in five features with dominant scores, namely TINN, alpha1, QRS width mean, QRS width std, Dq max. The five features have a mean and standard deviation value of 0.08 ± 0.02 and compared to 21 other features that have a mean and standard deviation value of 0.03 ± 0.02 . For testing on type gain, the dominant features are TINN, RR std, LF power, Alpha2, hq min. The feature has a mean and standard deviation value of 0.10 ± 0.04 and the other 21 features have a mean and standard deviation value of 0.02 ± 0.01 . Furthermore, the features selected using type cover are TINN, QRS amplitude std, LF power, alpha1, QRS amplitude mean. The score calculation produces a mean value and standard deviation of 0.07 ± 0.01 for the 5 selected features. The other 21 features have mean and standard deviation values of 0.03 ± 0.01 .

Based on the calculation of the mean and standard deviation values obtained, it is known that type gain has the best score calculation composition. Therefore, the use of type gain as an importance type method is very reasonable to be tested on models with the most optimal parameter settings. In addition, it is known that the TINN feature is the most frequently appearing features for all three types. This illustrates that TINN obtained from feature set 1 (fs1) as part of the calculation of heartrate variability has an important role in generating dominant features. In addition, features obtained from feature set 7(fs7), such as QRS morphologies and feature set 5(fs5), such as alpha have an important role in complementing the presence of dominant features in all three types due to their appearance in each selection process. Fig 7 shows the difference in data point distribution of apnea and non-apnea recordings for the eleven best features obtained using type weight. All these features contribute to the improved performance of XGBoost with very promising results.

3.2 Optimization Configuration of XGBoost to improve evaluation performance

In this process, to obtain high accuracy results, it is necessary to adjust the configuration of the proposed model. In this research, there are five scenarios tested with the consideration that the selected configuration is the result of optimization on several parameters with the best performance. Settings are made to maximize and determine the role of parameters and determine their effect on model performance. Based on the test, the max depth and learning rate parameters become the optimal recommendations for demonstrating regularization of gradient boosting. While other parameter settings use the default XGBoost algorithm settings, as shown in Table 4.

The scenario implementation and performance results of each model are as shown in Table 5. From the table, The Exp1 is the experiment using parameter setting based on optimization general setting parameters, learning setting parameters and command line setting parameters. The experiment used coverage with the best five feature set. The feature set are TINN, QRS amplitude std, LF power, alpha1, QRS amplitude mean. The performance of the classification process in scenario Exp1 is 84.4%, 85.7%, 82.9%, and 84.4% for each accuracy, specificity, sensitivity, and precision. Parameter settings in the XGBoost scenario Exp2 model use a gain type with the five best features, namely, TINN, RR std, LF power, Alpha2, hq min. The performance of the model resulted in accuracy, specificity, sensitivity, and precision of 89.6%, 92.5%, 86.5%, and 89.8%, respectively. Meanwhile, scenario Exp3 produces a performance measure of accuracy of 92.2%, specificity of 92.9%, sensitivity of 91.4%, and precision of 92.2% based on five selected features as a result of selection using weight type, namely TINN, QRS amplitude std, LF power, alpha1, QRS amplitude mean. Furthermore, in scenario Exp4 with the best number of features as many as 11 features can produce performance measures on the parameters of accuracy, specificity, sensitivity and precision of 93.5%, 93%, 94.1% and 93.5%, respectively. As expected, the best classification performance was obtained in Exp5 with accuracy, specificity, sensitivity and precision of 96.1%, 100%, 92.1% and 96.4%, respectively. Fig. 8 illustrates how XGBoost reaches its final decision from the decision tree constructed in the Exp5 scenario using all features. Testing the parameter-optimized XGBoost model on all features yields a very promising improvement in performance. Neither the large number of features nor the small, class-imbalanced sample of apnea and non-apnea recordings prevents XGBoost from outperforming the other classification methods, as shown in Table 6. It reports the results of several machine-learning techniques, evaluated to confirm that the model is robust to under- or over-sized, structured, and multidimensional datasets while remaining accurate and reliable. The techniques were compared according to their underlying classification strategies for apnea and non-apnea. They included SVM, which is based on a kernel (neural-network-like) technique, as well as AdaBoost and XGBoost, which are based on ensemble learning implementation and kNN which works using Instance based principles. Based on testing using the testing set, it gives promising results with its application to ensemble learning-based machine learning, namely XGBoost. These results prove that the dataset used is very suitable for ensemble-based learning models. This ensemble boosting technique works sequentially which changes a set of weak learners, combines and selects strong learners based on the error rate generated

during learning. Fig. 6 also shows the decision tree constructed with all features. Based on this tree, several features dominate the final decision, drawn from feature sets fs1, fs2, fs3, fs4, fs5, fs6, and fs7.

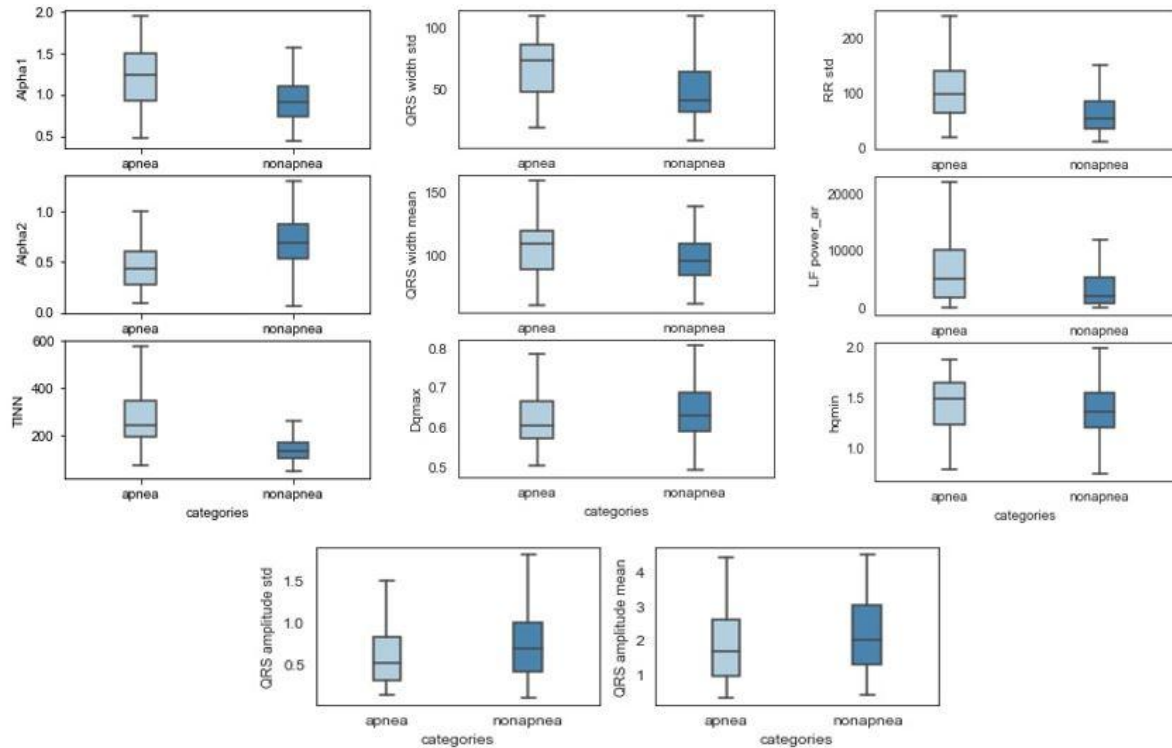


Fig. 7. Representative of boxplot for showing the variation of the best eleven importance features

TABLE 4.
OPTIMIZATION PARAMETER FOR VARIOUS XGBOOST MODEL

Classifiers	Optimization parameters
Optimization XGBoost model	base_score=0.5, colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1, importance_type='gain', learning_rate=0.5, max_delta_step=0, max_depth=4, min_child_weight=1, n_estimators=100, n_jobs=0, num_class=2, num_parallel_tree=1, objective='multi: softprob', random_state=0, reg_alpha=0, reg_lambda=1, silent=1, subsample=1

TABLE 5.
XGBOOST CLASSIFICATION RESULTS-BASED CONFUSION MATRIX WITH VARIOUS FEATURE SET AND MODEL OPTIMIZATION PARAMETERS

Scenario XGBoost Model	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Exp1: Optimization XGBoost model (coverage type)	84.4	85.7	82.9	84.4
Exp2: Optimization XGBoost model (gain type)	89.6	92.5	86.5	89.8
Exp3: Optimization XGBoost model (weight type)	92.2	92.9	91.4	92.2
Exp4: Optimization XGBoost model using weight type (the best eleven feature importances)	93.5	93	94.1	93.5
Exp5: Optimization XGBoost model using weight type (All features)	96.1	100	92.1	96.4

IV. DISCUSSION

This section discusses the experimental results. To our knowledge, the proposed framework is the first to provide a structured assessment of single-lead ECG-apnea based on multiple features, namely statistical measures of the RR interval, FFT- and AR-based spectral analysis, non-linear Poincaré plot analysis, entropy measures from mono-fractal and multifractal analysis, and QRS-morphology characteristics. The study aims to identify the features that carry the most clinically significant information; accordingly, the choice of an appropriate extraction method capable of capturing Apnea-ECG signal patterns is decisive. To assess the performance of the proposed framework, the best XGBoost model is evaluated using metrics such as accuracy, specificity, sensitivity and precision. The use of XGBoost in tests to classify sleep disorders is a useful and reliable choice when compared to other classification methods. On several occasions and

research results on different objects XGBoost was able to produce high accuracy and was the method of choice on several competition occasions (kaggle competition). In fact, XGBoost is able to outperform other adaboost family methods. The optimized XGBoost model is able to learn the training dataset maximally and evaluate it with excellent performance results resulting in accuracy, specificity, sensitivity and precision of 96.1%, 100%, 92.1% and 96.4% respectively on the testing dataset.

The XGBoost results are compared with prior work on the Apnea-ECG database in Table 7. The comparison considers the number of events per target class, the extraction method, the number of features, and the classification method, in order to assess how the method and the extracted features affect classification. The number of features is also examined because it influences the computational cost and is related to per-feature noise, redundant feature values, and irrelevant feature sets.

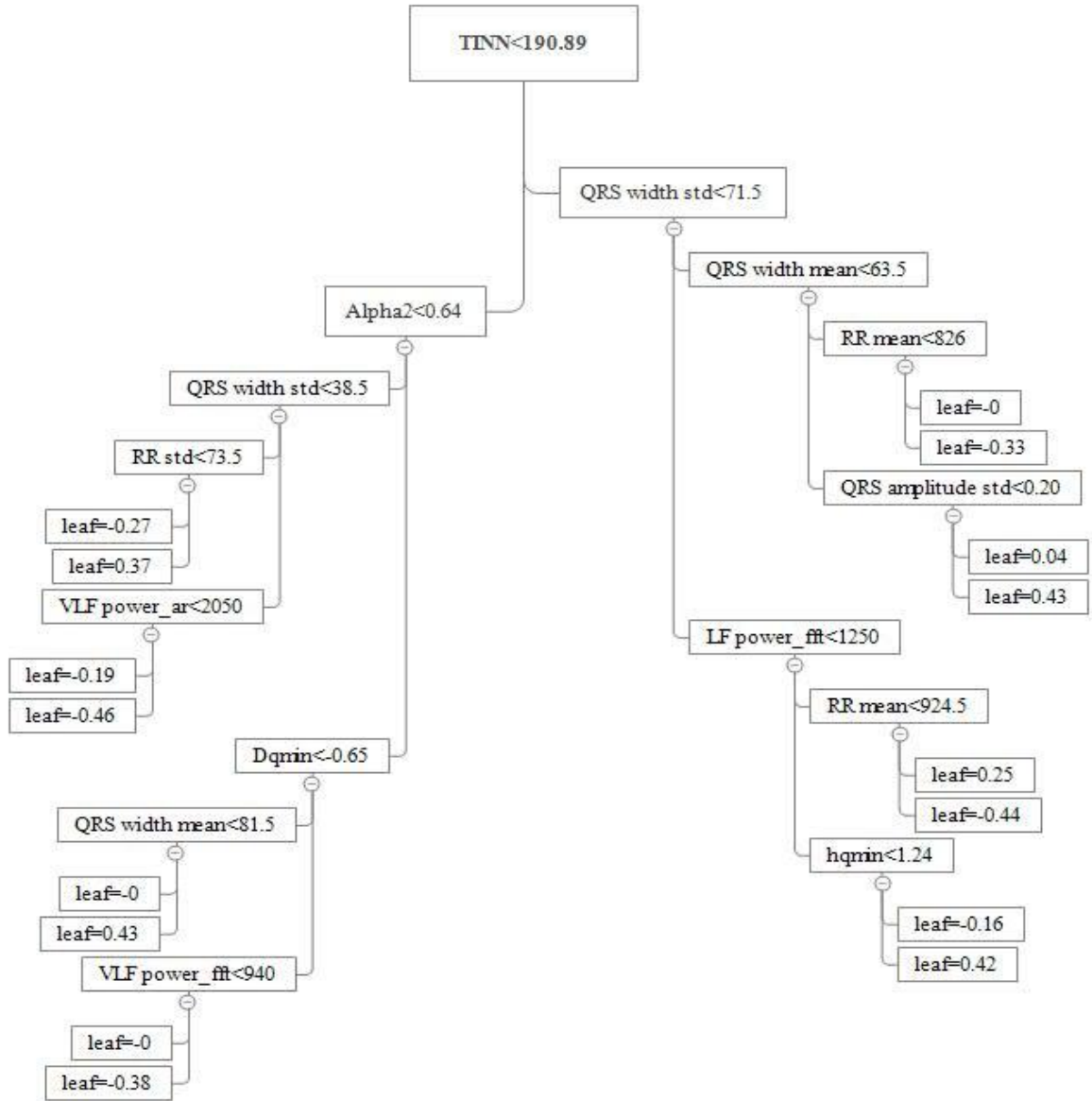


Fig. 8. The construction decision tree of the optimization XGBoost model (All features) to provide the final decision.

TABLE 6.
CLASSIFICATION PERFORMANCE RESULTS FOR VARIOUS METHODS

Various Classifiers	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Artificial Neural Network (neural networks based)	83.1	95.8	77.6	65.7
Gradient Boosting Classifier (ensemble based)	94.8	95.2	94.3	94.3
Random Forest Classifier (ensemble based)	93.5	94.1	93.0	91.4
kNN (instance based)	85.7	89.7	81.6	88.6
Decision Tree Classifier (decision Tree based)	87.0	82.0	91.1	91.4
SVM- RBF kernel (neural networks based)	80.5	75.5	91.7	62.9
SVM-Linear kernel (neural networks based)	83.1	83.7	82.4	80.0
SVM-Sigmoid kernel (neural networks based)	71.4	95.5	61.8	97.1
SVM-Polynomial kernel (neural networks based)	75.3	73.5	78.6	62.85
AdaBoost (ensemble based)	90.9	90.7	91.2	88.6
XGBoost (ensemble based)	96.1	100	92.1	96.4

TABLE 7.
CLASSIFICATION PERFORMANCE COMPARISON ON APNEA-ECG DATABASE

References	Methods	Classifiers	Number of features	Accuracy (%)	Specificity (%)	Sensitivity (%)
Nguyen[36]	Recurrence quantification analysis	soft decision fusion from SVM and NN	72	85.26	83.47	86.37
Hasan[37]	Tunable-Q factor wavelet transform (TQWT) domain	Adaptive Boosting	36	87.33	90.72	81.99
Hassan and Haque[38]	Tunable-Q factor wavelet transform (TQWT) domain	RUSBoost	3	88.88	87.58	91.49
Tripathy[39]	Energy and fuzzy entropy	Kernel extreme learning machine	12	-	74.64	78.02
Janbakhshi et al[19]	ECG-derived respiration (EDR) and RR intervals of ECG	ANN	85	90.9	91.8	89.6
Babaeizadeh et al[40]	Wavelet transform and non-linear entropy	SVM using RBF kernel	12	92.98	93.75	91.74
Fahrzi, Iman[41]	ECG morphologicals	RBF kernel function of SVM	17	85.1	-	-
F. Faccin et al [42]	Derived respiratory signals	hybrid deep model and Markov Chain	28	86	80	95.8
T. Yamane [43]	RR intervals and R-peak amplitudes	various machine learning	18	-	95.7 to 100	91.7 to 100
This work	Multi-feature non-linear analysis	XGBoost	26	96.1	100	96.4

These considerations underscore the importance of analyzing in detail the relationships within each feature group and combining them through approaches suited to the overall ECG signal and its QRS morphology. For example, recurrence quantification analysis has been applied to spectral time-series data derived from the heart-rate variability of ECG signals, treated as a non-stationary representation, to obtain recurrence plots [36]. The selection method eliminated 72 features that were not relevant to the target class, ultimately retaining 33 features given the large sample size.

In another study, an entropy-analysis method using 12 features achieved 94.63% accuracy with an RBF-kernel SVM on one-minute recordings, although the number of events in the training and testing sets was not clearly reported [40]. Decision trees have also been applied to sleep-disorder detection using adaptive boosting, which achieved 87.33% accuracy with 150 trees yielding the lowest error [36]. Another boosting-based study [38] used random under-sampling boosting and a tunable Q-factor wavelet transform to extract three input features. Each event lasted one minute (6000 samples), but the numbers of samples in the training and testing sets were not specified. These studies suggest that decision-tree-based modelling is a promising direction for improvement. In the present study, the contribution of XGBoost is evident: combining decision tree and ensemble techniques to reduce the error at each iteration from weak learners offers a practical aid for clinicians in classifying sleep disorders. Even with a limited and unequal number of events in the training and testing

sets, the results are promising, indicating that the complexity of the recordings and the number of events can yield higher accuracy when the entire pipeline dataset preparation, preprocessing, extraction, feature selection, and classification is carried out appropriately. As a result of this sequential process, a combination of methods well matched to the characteristics of the dataset maximizes the discriminability of the sleep-disorder signal patterns. The recognized patterns correlate with feature values suited to the machine-learning model in constructing the trees. From

the resulting decision tree, the features that influence the final decision can be identified, including TINN, RR std, and RR mean from feature set 1; LF power and VLF power (FFT) from feature set 2; VLF power (AR) from feature set 3; alpha1 and alpha2 from feature set 5; Dqmin and hqmin from feature set 6; and QRS width std, QRS width mean, and QRS amplitude mean from feature set 7. In spite of employing a range of signal processing strategies in electrocardiogram (ECG) analysis such as deriving respiratory signals, accurately calculating RR intervals, precisely detecting R-peaks, and conducting morphological evaluations of waveform shape and duration the overall performance is still lacking[41], [42]. These techniques, despite their technical precision, have failed to consistently differentiate between apnoea and non-apnea occurrences. Additionally, the amalgamation of varied feature sets has also been insufficient in reliably separating diseased from non-pathological circumstances, emphasising the demand for more advanced or integrative analytical approaches.

To assess the advantages of XGBoost over other classification methods on the same database, several classifiers were also evaluated using the scheme shown in Fig. 1, with the performance results reported in Table 7. The performance of any single tree is of limited significance, because the accuracy of an ensemble is determined by the combined contribution of all its trees, each of which mainly refines the predictions of the preceding trees in the sequence. As an effort to improve sleep-disorder detection, this study still has limitations. First, the test dataset is not completely free of noise, which can reduce the clinical information available if it is not attenuated or removed; such noise includes low- and high-frequency interference [44]. Secondly, each ECG recording was screened only as either a non-apnea or an apnea signal, whereas in some cases there may be an underlying potential for cardiac dysfunction or other heart diseases. These limitations therefore warrant further investigation to develop a decision-support system that can assist medical personnel in clinical decision-making.

V. CONCLUSIONS

This study demonstrates that the proposed approach achieves substantially higher accuracy than is typically attainable with existing boosting algorithms and other classifiers. XGBoost is a high-performance gradient-boosting framework based on decision trees and is widely used for classification and many other machine-learning tasks. It also performs well on large datasets and long-duration signals, although training can be time-consuming. In our experiments, the feature set combining RR-interval, QRS-morphology, and non-linear features achieved accuracy, specificity, and sensitivity of 96.1%, 100%, and 92.1%, respectively. These findings indicate that the XGBoost classifier is a competitive and powerful tool for automatic classification that can reduce human intervention in future medical diagnosis.

ACKNOWLEDGMENT

This research was supported by BIMA Research Grant by Ministry of Higher Education, Science, and Technology, Republic of Indonesia (Grant No: 059/C/DT.05.00/KPA/2025 and No: 167/C3/DT.05.00/PL-BATCH II/2025). Furthermore, this work was also supported by Artificial Intelligence for Healthcare and Biomedical Research Group, Center of Excellence TREEMINE, Politeknik Negeri Batam.

AUTHOR CONTRIBUTIONS

Conceptualization, Iman Fahruzi; software, Iman Fahruzi, Muhamad Haikal Akmal, Abdurahman Dwijotomo; validation, Iman Fahruzi, Abdurahman Dwijotomo; writing original draft preparation, Iman Fahruzi; writing review and editing, Abdurahman Dwijotomo; visualization, Iman Fahruzi; supervision, Iman Fahruzi, Abdurahman Dwijotomo; project administration, Muhamad Haikal Akmal, Ridwan; funding acquisition, Iman Fahruzi.

FUNDING

This research was supported by BIMA Research Grant by Ministry of Higher Education, Science, and Technology (Grant No: 059/C/DT.05.00/KPA/2025 and No: 167/C3/DT.05.00/PL-BATCH II/2025)

DATA AVAILABILITY

The data is publicly available in <https://phy.sionet.org/content/apnea-ecg/1.0.0/>.

DECLARATIONS

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Clinical trial number:
Not applicable

Competing interests
The authors declare no competing interests.

REFERENCES

- [1] M. Svensson, K. A. Franklin, J. Theorell-Haglöw, and E. Lindberg, "Daytime Sleepiness Relates to Snoring Independent of the Apnea-Hypopnea Index in Women From the General Population," *Chest*, vol. 134, no. 5, pp. 919–924, Nov. 2008, doi: 10.1378/chest.08-0847.
- [2] M. F. Akşahin, B. Oltu, and B. K. Karaca, "Sleep apnea detection using blood pressure signal," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, May 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404271.
- [3] An American Academy of Sleep Medicine Review, M. S. Wise, D. L. Arand, R. R. Auger, S. N. Brooks, and N. F. Watson, "Treatment of Narcolepsy and other Hypersomnias of Central Origin," *Sleep*, vol. 30, no. 12, pp. 1712–1727, Dec. 2007, doi: 10.1093/sleep/30.12.1712.
- [4] S. I. Isidoro, A. Salvaggio, A. Lo Bue, S. Romano, O. Marrone, and G. Insalaco, "Effect of obstructive sleep apnea diagnosis on health related quality of life," *Health and Quality of Life Outcomes*, vol. 13, no. 1, p. 68, May 2015, doi: 10.1186/s12955-015-0253-1.
- [5] R. N. Aurora *et al.*, "The Treatment of Central Sleep Apnea Syndromes in Adults: Practice Parameters with an Evidence-Based Literature Review and Meta-Analyses," *Sleep*, vol. 35, no. 1, pp. 17–40, Jan. 2012, doi: 10.5665/sleep.1580.
- [6] D. W. Hudgel, "Sleep Apnea Severity Classification — Revisited," *Sleep*, vol. 39, no. 5, pp. 1165–1166, May 2016, doi: 10.5665/sleep.5776.
- [7] M. Singh, P. Liao, S. Kobah, D. N. Wijesundera, C. Shapiro, and F. Chung, "Proportion of surgical patients with undiagnosed obstructive sleep apnoea," *British Journal of Anaesthesia*, vol. 110, no. 4, pp. 629–636, Apr. 2013, doi: 10.1093/bja/aes465.
- [8] S. Jones, "Sleep-disordered Breathing and Cardiovascular Disease: An Outcome-based Definition of Hypopneas," *Yearbook of Pulmonary Disease*, vol. 2009, pp. 284–285, Jan. 2009, doi: 10.1016/S8756-3452(08)79131-X.
- [9] S. G. Pearse and M. R. Cowie, "Sleep-disordered breathing in heart failure," *European Journal of Heart Failure*, vol. 18, no. 4, pp. 353–361, 2016, doi: 10.1002/ehf.492.
- [10] P. E. Peppard, T. Young, M. Palta, and J. Skatrud, "Prospective study of the association between sleep-disordered breathing and hypertension," *N Engl J Med*, vol. 342, no. 19, pp. 1378–1384, May 2000, doi: 10.1056/NEJM200005113421901.
- [11] M. Arzt, T. Young, L. Finn, J. B. Skatrud, and T. D. Bradley, "Association of Sleep-disordered Breathing and the Occurrence of Stroke," *American Journal of Respiratory and Critical Care Medicine*, vol. 172, no. 11, pp. 1447–1451, Dec. 2005, doi: 10.1164/rccm.200505-702OC.
- [12] K. A. Franklin, C. Sahlin, H. Stenlund, and E. Lindberg, "Sleep apnoea is a common occurrence in females," *European Respiratory Journal*, vol. 41, no. 3, pp. 610–615, Mar. 2013, doi: 10.1183/09031936.00212711.
- [13] K. A. Franklin and E. Lindberg, "Obstructive sleep apnea is a common disorder in the population—a review on the epidemiology of sleep apnea," *Journal of Thoracic Disease*, vol. 7, no. 8, p. 12, 2015.
- [14] C.-C. Hsu and P.-T. Shih, "A novel sleep apnea detection system in electroencephalogram using frequency variation," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6014–6024, May 2011, doi: 10.1016/j.eswa.2010.11.019.
- [15] S. Chatterjee, S. Pratiher, and R. Bose, "Multifractal detrended fluctuation analysis based novel feature extraction technique for automated detection of focal and non-focal electroencephalogram signals," *IET Science, Measurement Technology*, vol. 11, no. 8, pp. 1014–1021, 2017, doi: 10.1049/iet-smt.2017.0117.
- [16] S. Babaeizadeh, D. P. White, S. D. Pittman, and S. H. Zhou, "Automatic detection and quantification of sleep apnea using heart rate variability," *Journal of Electrocardiology*, vol. 43, no. 6, pp. 535–541, Nov. 2010, doi: 10.1016/j.jelectrocard.2010.07.003.
- [17] H. Sharma and K. K. Sharma, "An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions," *Computers in Biology and Medicine*, vol. 77, pp. 116–124, Oct. 2016, doi: 10.1016/j.compbiomed.2016.08.012.
- [18] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, and S. V. Huffel, "A Novel Algorithm for the Automatic Detection of Sleep Apnea From Single-Lead ECG," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 9, pp. 2269–2278, Sep. 2015, doi: 10.1109/TBME.2015.2422378.
- [19] P. Janbakhshi and M. B. Shamsollahi, "Sleep Apnea Detection from Single-Lead ECG Using Features Based on ECG-Derived Respiration (EDR) Signals," *IRBM*, vol. 39, no. 3, pp. 206–218, Jun. 2018, doi: 10.1016/j.irbm.2018.03.002.
- [20] S. Martín-González, J. L. Navarro-Mesa, G. Juliá-Serdá, J. F. Kraemer, N. Wessel, and A. G. Ravelo-García, "Heart rate variability feature selection in the presence of sleep apnea: An expert system for the characterization and detection of the disorder," *Computers in Biology and Medicine*, vol. 91, pp. 47–58, Dec. 2017, doi: 10.1016/j.compbiomed.2017.10.004.
- [21] J. Gubbi, A. Khandoker, and M. Palaniswami, "Classification of sleep apnea types using wavelet packet analysis of short-term ECG signals," *Journal of Clinical Monitoring and Computing*, vol. 26, no. 1, pp. 1–11, Feb. 2012, doi: 10.1007/s10877-011-9323-z.
- [22] V. M. Panchagnula, V. Venkataiah, G. S. Murthy, A. Lakshmi, Y. Tondep, and V. S. Lakshmi, "Deep Learning Hybrid Models for ECG-Based Biometric Authentication and Health Forecasting in Telemedicine Systems," vol. 52, no. 11, 2025.
- [23] B. Guijarro-Berdiñas, E. Hernández-Pereira, and D. Peteiro-Barral, "A mixture of experts for classifying sleep apneas," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7084–7092, Jun. 2012, doi: 10.1016/j.eswa.2012.01.037.
- [24] K. Li, W. Pan, Y. Li, Q. Jiang, and G. Liu, "A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal," *Neurocomputing*, vol. 294, pp. 94–101, Jun. 2018, doi: 10.1016/j.neucom.2018.03.011.
- [25] V. Ho, C. M. Crainiceanu, N. M. Punjabi, S. Redline, and D. J. Gottlieb, "Calibration Model for Apnea-Hypopnea Indices: Impact of Alternative Criteria for Hypopneas," *Sleep*, vol. 38, no. 12, pp. 1887–1892, Dec. 2015, doi: 10.5665/sleep.5234.
- [26] U. J. Magalang *et al.*, "Agreement in the Scoring of Respiratory Events and Sleep Among International Sleep Centers," *Sleep*, vol. 36, no. 4, pp. 591–596, Apr. 2013, doi: 10.5665/sleep.2552.
- [27] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The apnea-ECG database," in *Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163)*, Cambridge, MA, USA: IEEE, 2000, pp. 255–258, doi: 10.1109/CIC.2000.898505.

- [28] T. Penzel, J. W. Kantelhardt, L. Grote, J. H. Peter, and A. Bunde, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 10, pp. 1143–1151, Oct. 2003, doi: 10.1109/TBME.2003.817636.
- [29] E. Ihlen, "Introduction to Multifractal Detrended Fluctuation Analysis in Matlab," *Frontiers in Physiology*, vol. 3, p. 141, 2012, doi: 10.3389/fphys.2012.00141.
- [30] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A: Statistical Mechanics and its Applications*, vol. 316, no. 1, pp. 87–114, Dec. 2002, doi: 10.1016/S0378-4371(02)01383-3.
- [31] M. Elgendi, B. Eskofier, S. Dokos, and D. Abbott, "Revisiting QRS Detection Methodologies for Portable, Wearable, Battery-Operated, and Wireless ECG Systems," *PLoS ONE*, vol. 9, no. 1, p. e84018, Jan. 2014, doi: 10.1371/journal.pone.0084018.
- [32] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.
- [33] H. Abrishami, M. Campbell, C. Han, R. Czosek, and X. Zhou, "P-QRS-T localization in ECG using deep learning," in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, Mar. 2018, pp. 210–213. doi: 10.1109/BHI.2018.8333406.
- [34] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785.
- [35] Y. Liang *et al.*, "Product Marketing Prediction Based on XGboost and LightGBM Algorithm," in *Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition*, in AIPR '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 150–153. doi: 10.1145/3357254.3357290.
- [36] H. D. Nguyen, B. A. Wilkins, Q. Cheng, and B. A. Benjamin, "An Online Sleep Apnea Detection Method Based on Recurrence Quantification Analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1285–1293, Jul. 2014, doi: 10.1109/JBHI.2013.2292928.
- [37] A. R. Hassan, "Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting," *Biomedical Signal Processing and Control*, vol. 29, pp. 22–30, Aug. 2016, doi: 10.1016/j.bspc.2016.05.009.
- [38] A. R. Hassan and Md. A. Haque, "An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting," *Neurocomputing*, vol. 235, pp. 122–130, Apr. 2017, doi: 10.1016/j.neucom.2016.12.062.
- [39] R. K. Tripathy, "Application of intrinsic band function technique for automated detection of sleep apnea using HRV and EDR signals," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 1, pp. 136–144, 2018, doi: 10.1016/j.bbe.2017.11.003.
- [40] A. Zarei and B. M. Asl, "Automatic Detection of Obstructive Sleep Apnea Using Wavelet Transform and Entropy-Based Features From Single-Lead ECG Signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1011–1021, May 2019, doi: 10.1109/JBHI.2018.2842919.
- [41] I. Fahruzi, "Towards Improvement of Sleep Apnea Identification using Multi-feature Analysis from Screening of Non-overlapping Single Lead ECG," vol. 51, no. 3, 2024.
- [42] F. Faccin, E.-H. Djermoune, P. Guyot, and L. Bougrain, "A Two-Stage Cascaded Ensemble based on CRNN and Markov Chain for Sleep Apnea Detection using ECG".
- [43] T. Yamane, M. Fujii, and M. Morita, "Clinical-level screening of sleep apnea syndrome with single-lead ECG alone is achievable using machine learning with appropriate time windows," *Sleep Breath*, vol. 29, no. 2, p. 156, Apr. 2025, doi: 10.1007/s11325-025-03316-0.
- [44] A. Kaur, A. Agarwal, R. Agarwal, and S. Kumar, "A Novel Approach to ECG R-Peak Detection," *Arabian Journal for Science and Engineering*, vol. 44, no. 8, pp. 6679–6691, Aug. 2019, doi: 10.1007/s13369-018-3557-8.