SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V13I2P109

Volume 13 Issue 2, 118-133, February 2026

Original Article

An Insomnia Detection Model using Augmented Two-Class Multichannel EEG Frequencies

Steffi Philip Mulamoottil[1], T. Vigneswaran[2]

[1,2]*School of Electronics Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India.*

[2]*Corresponding Author : vigneswaran.t@vit.ac.in*

Received: 16 November 2025 Revised: 20 December 2025 Accepted: 17 January 2026 Published: 17 February 2026

***Abstract** - Clinical diagnosis of Insomnia relies on sleep-stage information from multiple electrodes for an accurate prediction. Existing insomnia detection models achieved satisfactory performance with single-channel training samples based on sleep-stage epochs, but the performance of classifiers using multiple-channel information has not been investigated. Also, differences in sleep stage distribution might lead to non-uniformity across training subsets. This study develops a model that does not rely on sleep-stage analysis, instead using traditional signal augmentation to efficiently manage multichannel data through time-frequency analysis of mixed-frequency characteristics across various Electroencephalogram (EEG) channels. A two-class Full-Sleep Ensemble Learning (FS-EL) model is implemented using augmented dual-frequency sub-bands that characterize deep sleep across EEG frequencies. It builds new sub-band representations using frequency cropping and superposes the components in subsequent iterations, reformulating them into a two-class model. The features extracted from newly generated composite signals have been evaluated using an Ensemble Bagged Decision Tree (EBDT), Random Forests (RF), and Gradient Boosting (GB). It achieves superior performance with a composite two-class FS-EL model trained across various datasets on Out-Of-Bag (OOB) samples. FS-EL model attained classification accuracy and Area Under The Curve (AUC) of 0.99 using k-fold cross-validation on EBDT, and sensitivity of 1.0 compared to composite five-class and non-composite single-class models. The proposed model improves performance by generating new composite signals that exploit multichannel information using a subset of the raw data, achieving perfect distinction of data points and obtaining efficient sample classification.*

***Keywords** - Cross-Validation, Ensemble Models, Machine Learning, EEG Sub-Band Processing, Time-Frequency Analysis, Wavelet transform.*

1. Introduction

Sleep disorders use polysomnographic information to develop classification models, specifically, overnight sleep recordings of physiological signals for insomnia diagnosis. On average, 50-70 million people in America suffer from sleep disorders, where Insomnia, sleep apnea, narcolepsy, etc., are found to be common [1-3]. Insomniacs exhibit disturbed sleep [4], prolonged sleep latency, abnormal nocturnal awakenings, and so on [5]. Sleep disorder diagnosis can be performed using a Polysomnographic (PSG) study, which comprises biological signals such as Electroencephalogram (EEG), Electrocardiogram (ECG), Electrooculogram (EOG), Electromyogram (EMG), and so on [6]. Sleep disorder detection involves spectrograms of small signal segments with automated feature extraction using Neural Networks for Deep Learning, or Hand-Crafted features for Machine Learning (ML) classification [7-10]. Most supervised learning models for sleep disorder detection use either sleep stage information or single-channel data from physiological signals in PSG to build the models. Besides Polysomnographic (PSG) analysis, actigraphy data can also be used to assess sleep parameters and yield significant classification results in research studies [11-13], but it has drawbacks and shows lower performance. Among the physiological signals in PSG, EEG (amplitude range: 100-500 μV) is considered the gold standard and is recommended for developing insomnia-detection models [14]. It shows brain activity during sleep and its characteristic variations. Insomniacs exhibit deviations in sleep patterns across sleep stages and EEG frequencies compared with healthy subjects. Investigating abnormal pattern differentiation in time, frequency, and time-frequency domains involves filtering relevant frequency components, removing artifacts, and extracting features for diagnosis [15-18].

1.1. Related Works

There was limited research on automatic identification of Insomnia using supervised or unsupervised learning models. Sharma M et al. [19] developed a classification model using a triplet-half-band filter bank and wavelet decomposition with two EEG channels, achieving accuracies of 91.25% to 99.23%. A 19-channel sleep EEG was conducted to investigate the Neurophysiological features of Sleep Apnea, Analyzing Spectral Power, Network, and Microstate characteristics, with an accuracy of 88.3%, sensitivity of 92%, and specificity of 84% [20]. Additionally, the multimodality of biological signals in sleep staging was examined using an AdaBoost-random forest machine learning classifier that used three signals, achieving accuracies ranging from 92.25% to 96.83% on the CAP sleep dataset [21]. Besides Machine Learning, a three-dimensional Convolutional Neural Network (CNN) used multichannel EEG signals for automatic seizure detection, converting time-series signals into two-dimensional images, achieving 90% accuracy, 88.90% sensitivity, and 93.78% specificity [22]. EEG-based models utilize time-series segmented epochs based on sleep annotations available in public databases, yielding significant performances [23-28]. Similarly, insomnia identification models based on ECG and EOG were also seen to be superior in their accuracy [29-32]. In addition to PSG signals, actigraphy data are used to develop automated models for detecting insomnia [33-35].

Research studies on insomnia detection models using biological signals did not leverage the characteristics of multichannel signal analysis in the time and frequency domains. However, time and frequency-domain characteristics can be explored for sleep apnea detection by converting ECG signals into scalograms [36]; this approach also uses single-lead ECG signals. Likewise, sleep Apnea detection performed using a CNN-Transformer-LSTM model also uses single-lead ECG signals [37]. In another study on sleep disorder detection, Bruxism used two EEG channels for decision-tree classification [38], combining the channels using the Welch method for power spectral density feature extraction. A C3 and C4 EEG recording was used in [39], collected from subjects with paradoxical Insomnia, where singular spectra were computed from EEG segments and fed to an Artificial Neural Network (ANN) for EEG classification. From the literature, it is observed that most sleep disorder detection models used single-channel recordings for classification, or if multiple channels were used, time and frequency-domain characteristics were not extracted and provided to the classification models.

The study presented here addresses challenges posed by sleep stages, including the non-uniformity of segmented epochs within them, by using a subset of the input EEG. It aims to develop a model independent of manual interference in decision-making. Baseline insomnia detection models used short, segmented epochs based on manual sleep-stage annotations of single-channel EEG signals provided by a technician during recording, followed by classification using feature extraction from small time segments. In this case, it is challenging to manage multichannel information as used in clinical diagnostic interpretation. Moreover, features derived from sleep stages also exhibit an imbalance in the training data. Hence, a novel multichannel EEG detection model is presented in this paper that captures the time-frequency distribution of EEG sub-bands simultaneously across multiple channels. The study uses new signals from an existing dataset with sufficient training data via frequency cropping. It follows the principle of superposition for a detailed analysis, which gives diversity into the training data without altering the input characteristics. Hence, the study develops a simpler classification model that shows a diverse nature of EEG signals in the time-frequency domain.

The highlights of the proposed work are:

- The proposed approach develops a novel two-class FS-EL insomnia model using augmented signals of multichannel EEG sub-band frequencies. It tries to build a model independent of sleep stages for insomnia detection, leveraging characteristic variations of sleep EEG between insomniacs and healthy subjects.
- The model utilizes time and frequency analysis of sleep EEG from multiple channels with reduced complexity and generates data subsets using the statistical and entropy features, and individual sub-band power, extracted from newly created synthetic signals.
- The proposed two-class model extracts features from augmented two-class signals and achieves both accuracy and AUC of 0.99, and a sensitivity of 1.0 using an FS-EL model, bringing a few evaluation scores ahead in comparison to the two-class Light Sleep EL (LS-EL) Model, Traditional Composite Five-Class Model, and Non-Composite Single-Class Model.
- The two-class augmented data were trained on ensemble models such as EBDT, RF, and GB, where the proposed FS-EL model using EBDT, utilizing OOB samples, obtained a more satisfactory classifier performance. The results were re-evaluated using k-fold and stratified k-fold validation.

2. Materials and Methods

EEG datasets yielded improved classification performance with either spectrograms or scalograms, and with deep or machine learning classification based on features extracted from segmented epochs of physiological signals. Existing models use single-channel data comprising five sleep-stage sub-classes for classifier training. The baseline detection models have never leveraged the time- and frequency-domain properties of compound EEG to classify Insomnia from healthy subjects beyond sleep-stage analysis. The variation in EEG patterns across multiple electrodes plays a vital role in decision-making during clinical diagnosis, yet this approach has not been incorporated into modern detection models. Among numerous related works, it is noted that short segments of EEG frequency bands have not yet achieved sufficient prevalence in creating data classes for analytical purposes; instead, they use smaller time segments based on sleep stages. Including all sleep stages from multiple channels for the prediction process increases model complexity and disrupts the balance of the training subsets used by classification algorithms. Hence, the work proposes a novel approach that uses only a subgroup of raw data that leverages the composite nature of multichannel EEG by transforming the input into intermediate EEG sub-bands using the discrete wavelet transform, thereby finally creating the newly generated composite signals using only significant frequencies that highly differentiate insomniacs from healthy, leading to the development of a two-class model. The occurrence of low-frequency sleep waves, a marker of deep sleep, is reduced in individuals with Insomnia. Hence, those frequency ranges can help the classifiers distinguish data points more efficiently than using all sleep frequencies. The developed model was then compared for precision using five-class and single-class data grouping.

2.1. Proposed Approach

The current study uses the Cyclic Alternating Pattern (CAP) sleep database from the PhysioNet website [40, 41], which comprises PSG recordings representing various sleep disorders. The sleep insomnia recordings were found to have different sampling frequencies; therefore, to maintain uniformity, only signals with a sampling frequency of 512 Hz were selected for the proposed work. Based on the availability, data from C4-A1, FP2-F4, F4-C4, C4-P4, and P4-O2 were chosen to develop a multichannel detection model. The sleep brain activities, recorded from different channels, are shown in Figure 1. A sleep EEG is a temporal aggregation of neuronal responses, maintained in spatial alignment, that displays the brain's activity at different frequencies, such as delta, theta, alpha, sigma, and beta, which are visible in the ranges 0.5-4 Hz, 4-7 Hz, 8-12 Hz, 12-16 Hz, and 13-30 Hz, respectively.

Individuals with sleep disorders, such as Insomnia, have a direct effect on EEG frequencies during sleep. The healthy and insomniac nature of sleep varies across frequencies, which helps classify the given data. The mixed behavior of EEG frequency components is a key concept for the study presented here. The baseline models segmented EEG epochs according to the sleep stages specified in the sleep stage annotations to build a classification model. The proposed work, illustrated in Figure 2, presents a simplified layout of a classification model that uses only frequencies that indicate significant sleep states, efficiently differentiates input data points, and achieves accurate classification, thereby reflecting the composite nature of EEG signal frequencies. The work uses raw EEG signals from five channels, which were filtered with a Butterworth band-pass filter, and the resulting filtered signals spanning 0.5-45 Hz were cropped into their constituent frequencies. The Signal-To-Noise Ratio (SNR) of the chosen channels indicates the quality of the EEG signals, as shown in Table 1; higher SNR values indicate a clearer EEG signal.

Table 1. SNR of selected channels after filtering

Channel	SNR
C4-A1	42.1558
FP2-F4	43.7163
F4-C4	46.6930
C4-P4	50.8452
P4-O2	42.6667

The cropped sub-bands bearing a single frequency were obtained using the discrete wavelet transform, and are selectively used to create new composite signals that define complete and partial sleep stages. Henceforth, a detailed time-frequency-domain analysis can be performed using shorter time segments of the newly generated signals, which are then fed to the feature extraction unit for sub-band power, statistical, and entropy features. Further, the extracted data are fed into the FS-EL classification model, which comprises only low-frequency features that are more significant for differentiating insomniacs from healthy individuals, while neglecting the remaining ones, and performs cross-validation to generalize to new, unseen data. In the presented model, the study has used the principle of superposition as a basis for outlining sleep states and for leveraging the composite nature of EEG signals for insomnia detection. It generates composite signals from raw EEG across multiple channels that highlight insomniacs' behavior during sleep-stage transitions.

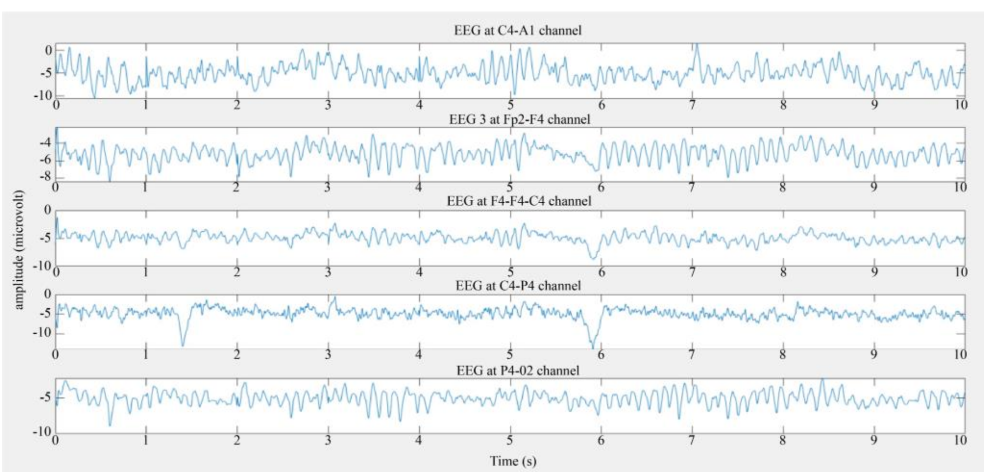


Fig. 1 EEG pattern at different channels

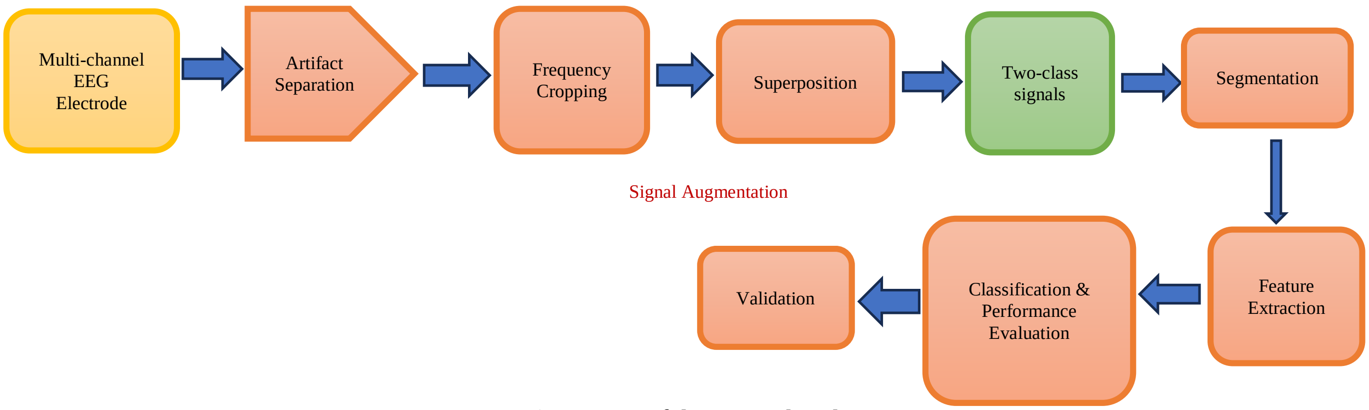


Fig. 2 Layout of the proposed work

It then segments the generated composite signals by sleep-stage duration, e.g., 30-second epochs, thereby performing time-frequency analysis with the novel two-class model and comparing the results with those from conventional composite and non-composite models.

Among the coefficients of filtered signals decomposed using the Discrete Wavelet Transform (DB8), low-frequency detail and approximation coefficients were chosen to generate superimposed brain patterns for the FS-EL model.

2.1.1. Two-Class Augmented Sub-Band Generation

The human brain can be considered a linear system, in which brain activity can be recorded using EEG oscillations with sleep stages referenced to five significant EEG frequencies.

Conventionally, EEG frequencies overlap and can be decomposed into various partitions using the Fast Fourier Transform (FFT), the Short-Time Fourier Transform (STFT), or the Wavelet Transform (WT) [42, 43], thereby diminishing the compounded characteristics of signals.

On the other hand, the superposition of EEG frequencies can be utilized for Neurological diagnostic purposes [44]. In the proposed work, a natural blend of EEG frequencies resembling clinical sleep-stage variations in insomniacs is used to generate subsets via superposition, which is illustrated in Figure 3. The brain's response during sleep is equivalent to the summation of different brain frequencies' potential or various electrical oscillations embedded in it.

The choice of frequency selection for generating a composite EEG sub-band should reflect the clinical diagnostic nature of Insomnia. The augmented two-class composite signals and decomposed frequencies from the source waveform of a single channel are depicted in Figure 4. Likewise, the augmentation repeats for the remaining chosen channels.

The new composite signals for LS-EL and FS-EL models correspond to high-frequency and low-frequency sleep states, respectively. Hence, the five-class EEG structure for insomnia detection was reduced to a two-class scheme for precision and simplicity, comprising two detection models based on superimposed high- and low-frequency synthetic signals.

Both LS-EL and FS-EL models follow a two-class scheme; hence, to determine the best choice among the models that use composite signals, the study identified the one that yields the highest classification performance.

An EEG can be viewed as a composite signal composed of multiple frequency components. In the proposed model, a new composite signal is reconstructed from the independent EEG frequency components across various channels.

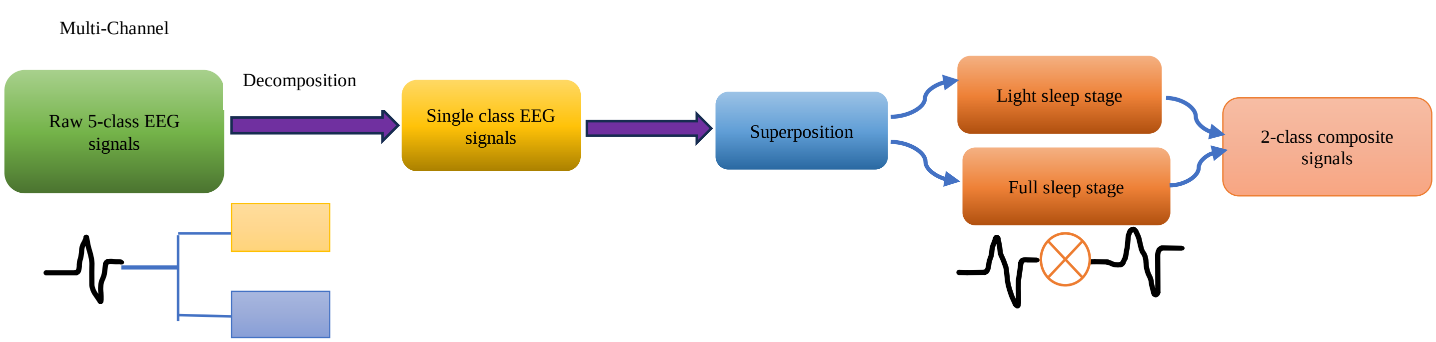
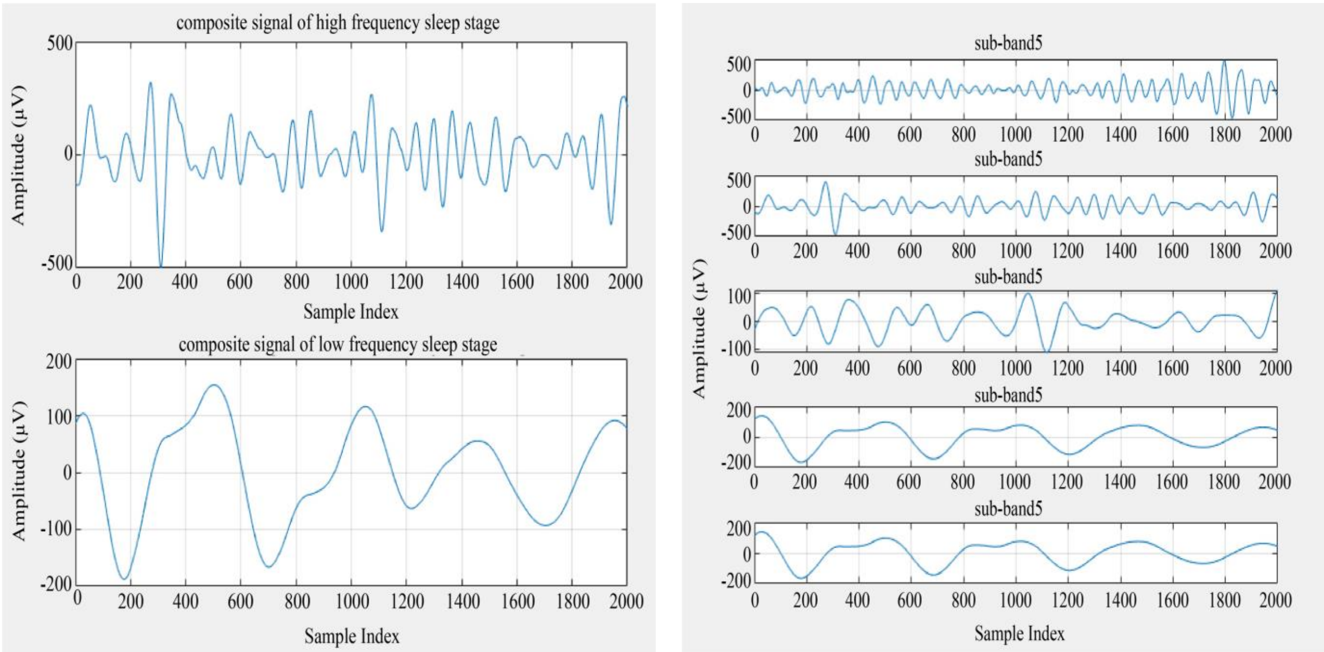
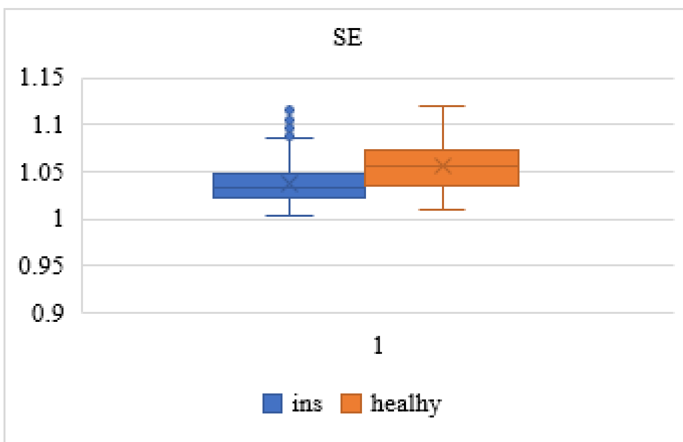


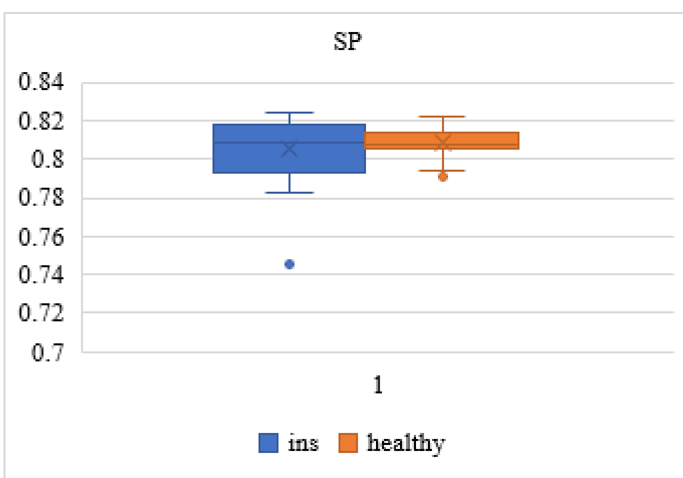
Fig. 3 Two-class composite signal generation

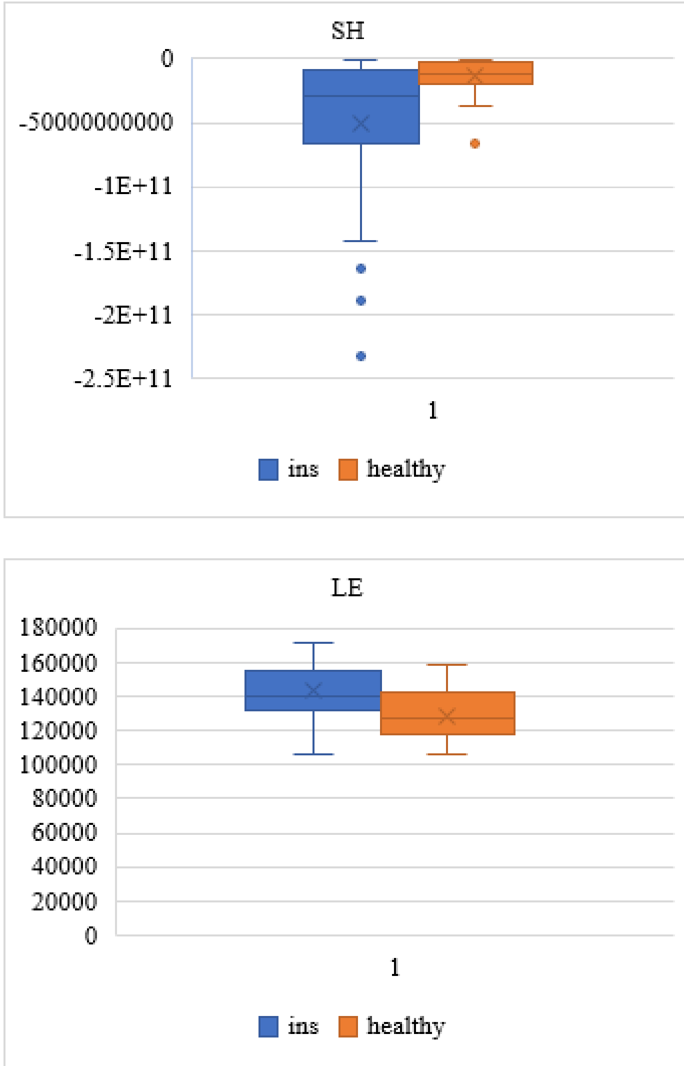


(a) **(b)**

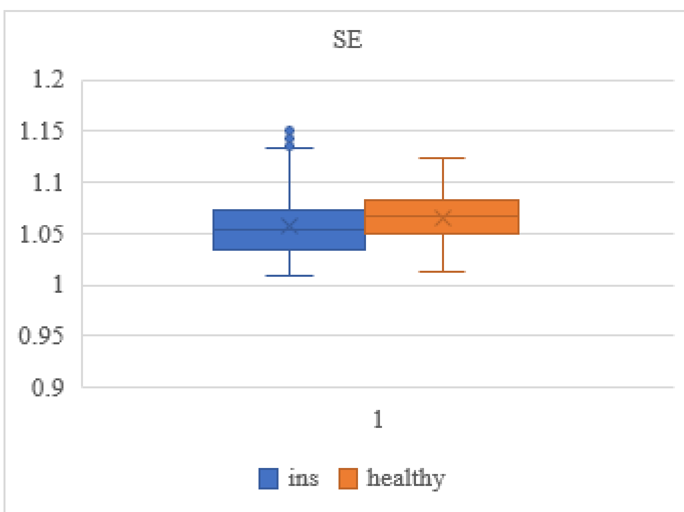
Fig. 4 The plot of newly generated composite sub-bands vs Individual sub-bands, (a) Composite EEG two-frequency sub-bands for high frequency LS-EL and low frequency Fs-EL models, and (b) Decomposed EEG single-frequency sub-band.

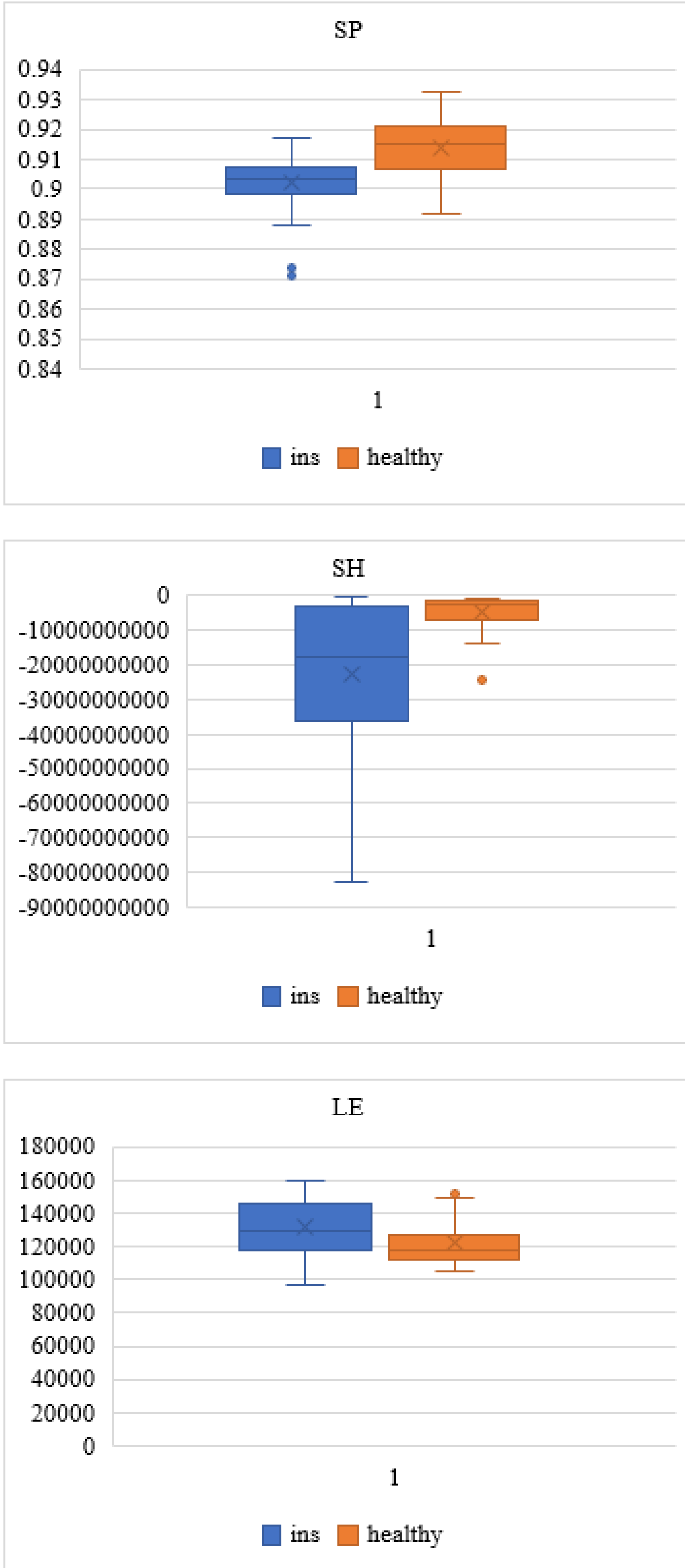




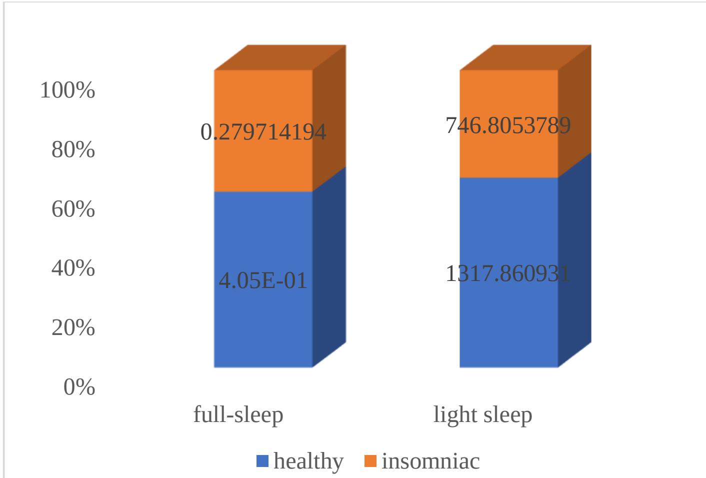

SH
0
-50000000000
-1E+11
-1.5E+11
-2E+11
-2.5E+11
1
ins
healthy
LE
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
1
ins
healthy

(a)

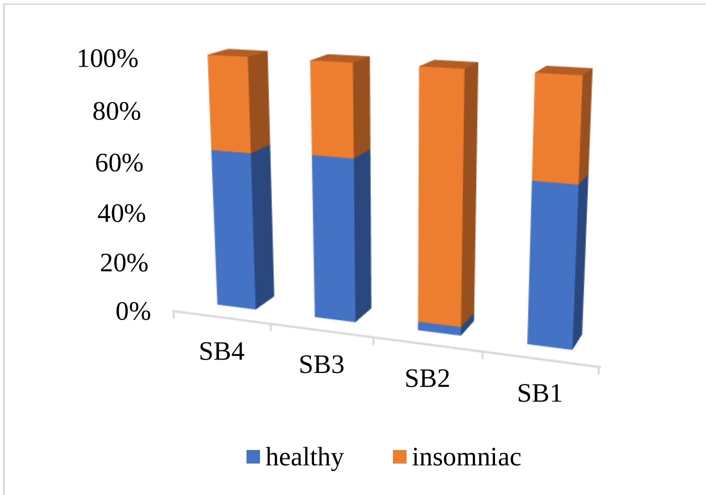

SE
1.2
1.15
1.1
1.05
1
0.95
0.9
1
ins
healthy


SP
0.94
0.93
0.92
0.91
0.9
0.89
0.88
0.87
0.86
0.85
0.84
1
ins
healthy
SH
0
-10000000000
-20000000000
-30000000000
-40000000000
-50000000000
-60000000000
-70000000000
-80000000000
-90000000000
1
ins
healthy
LE
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
1
ins
healthy

(b)



(c)



(d)

Fig. 5 Boxplots of the features from (a) Full-sleep, (b) Light-sleep, (c) Band power of superimposed frequencies, and (d) Band power of individual sub-band frequencies. Ins: features extracted from insomnia subjects, healthy: healthy subjects, and sb: indicates sub-band. Circles outside the box indicate the outliers in the extracted feature distribution.

The highest frequencies are excluded from composite signal creation because they play a minimal role in reflecting sleep variations. The feature extraction unit derived statistical features, including mean, standard deviation, skewness, and kurtosis. The entropy features such as the Sample Entropy (SE), Spectral Entropy (SP), Shannon Entropy (SH), And Log Energy Entropy (LE), of both LS-EL and FS-EL stages and the band power as a function of superimposed frequencies and single-frequency sub-bands, as shown in Figure 5 (a), (b), (c), and (d), is essential for developing Machine Learning Models, indicating a clear distinction between insomniacs and healthy individuals, making classification efficient using ensemble models.

2.1.2. Classification

In the proposed two-class FS-EL model, the features extracted from the newly generated signals were fed into the classifier. The data subset was split into training and test sets at an 80:20 ratio, with OOB samples from the train set used for prediction to reduce overfitting, improve generalizability, and better formulate the model for test data. An EBDT is a group of weak decision tree learners combined into a single structure where predictions are made by majority voting across the outputs of each tree [45, 46].

Ensemble bagging improves the model's performance and reduces overfitting compared to single decision tree models. As decision trees experience high variance, the ensemble bagged model exhibits low variance. The FS-EL EBDT model is used with OOB samples, as shown in Figure 6. The model's performance was cross-validated using 5-fold cross-validation, though this is considered an additional validation step, even if OOB samples serve the same purpose. The two-class classification was also implemented using high-sleep-frequency segments, which correspond to Light Sleep (LS) in the EEG signal.

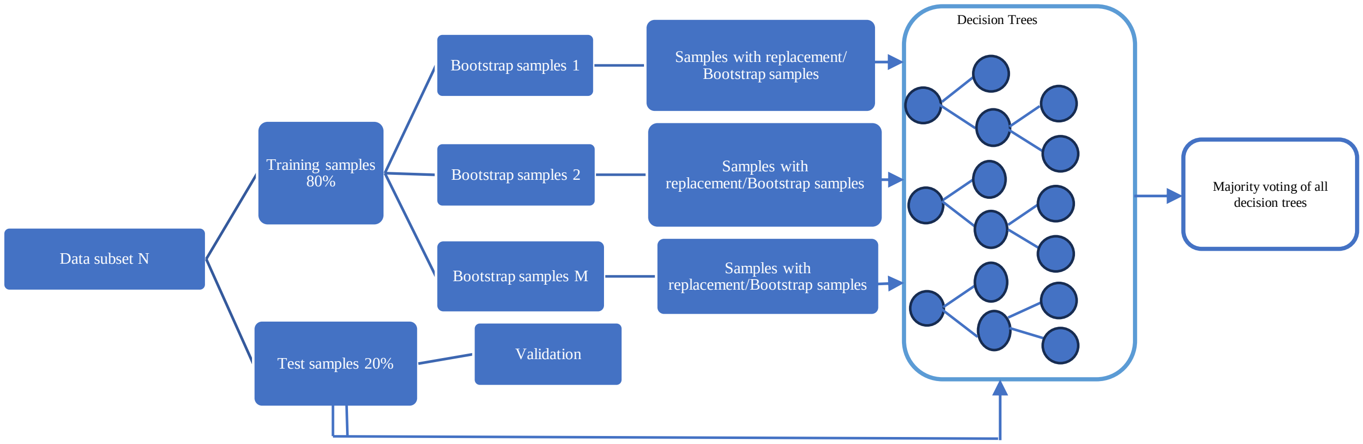


Fig. 6 FS-EL EBDT model for classification of two-class composite signals

The new LS-EL model also uses OOB samples for classification to determine whether the former or latter plays a significant role in improving classifier performance. The two-class models were also evaluated using two other ensemble learning methods, namely RF and GB, to determine which method is better suited for two-class signal classification. When building RF, the decision trees split in such a way that a random subset of predictors is considered from the entire set of predictors, where if m is the number of predictors selected for the split from the p set of predictors, then m will be approximately equal to the square root of p [47].

Random forests work by taking a random subset of samples and predictors [48], unlike bagging. The GB is an improved version of AdaBoost that uses gradient descent to optimize loss functions and generate new decision tree models [49]. The enhanced EBDT model illustrated in Figure 6 uses the data sub-sets extracted from high and low-frequency two-class composite signals.

3. Results and Discussions

3.1. Results

The model creates two-class composite signals from multichannel EEG sub-band frequencies, thereby reducing the complexity of using EEG frequencies across multiple channel electrodes. The prediction of two-class signals has been experimented with using ensemble learning algorithms, including EBDT, RF, and GB. The experiment has been conducted by transforming a five-class EEG into a mixture of two-class signals using subsequent frequencies across multiple EEG channels. Hence, low-frequency EEG components form the first composite superimposed signal (the low-frequency full-sleep signal), and the high-frequency signals that transition a person from light to deep sleep are called high-frequency light-sleep signals. The accuracy of the EBDT classifier in distinguishing between light and full sleep, achieved through superposition, is illustrated in Figure 7.

The classification results for light and full sleep characteristics, obtained by combining light and full sleep features after the feature extraction stage without using superposition or composite signals, are also shown in Figure 7. SP indicates a superposed signal, and clean represents the generation of training subsets by combining the features obtained from the corresponding individual sub-bands without applying superposition.

The FS-EL model with full sleep low-frequency components yields higher classification accuracy than the other scenarios. It also highlights the importance of combining low-frequency components based on superposition to generate new signals. The detailed metrics for the light and full sleep two-class signals, showing the model performance across different classifiers, are depicted in Table 2. Table 3 presents the classifier's performance using features extracted from non-composite two-class signals.

Table 2. Classifier performance for two-class composite signals

Performance Metrics	Full-sleep model			Light-sleep model		
	EBDT	**RF**	**GB**	**EBDT**	**RF**	**GB**
Accuracy	0.984	0.965	0.965	0.950	0.934	0.94
Sensitivity	1.0	1.0	1.0	0.988	0.96	0.96
Specificity	0.97	0.935	0.935	0.916	0.90	0.92
Precision	0.968	0.930	0.930	0.911	0.90	0.914
F1-score	0.983	0.964	0.964	0.948	0.93	0.937
AUC	0.99	0.989	0.997	0.986	0.99	0.986

Table 3. Classifier performance for non-composite two-class signals using EBDT

Metrics	Full sleep model	Light sleep model
Accuracy	0.913	0.884
Sensitivity	0.937	0.95
Specificity	0.89	0.81
Precision	0.89	0.838
F1-Score	0.915	0.89
AUC	0.98	0.958

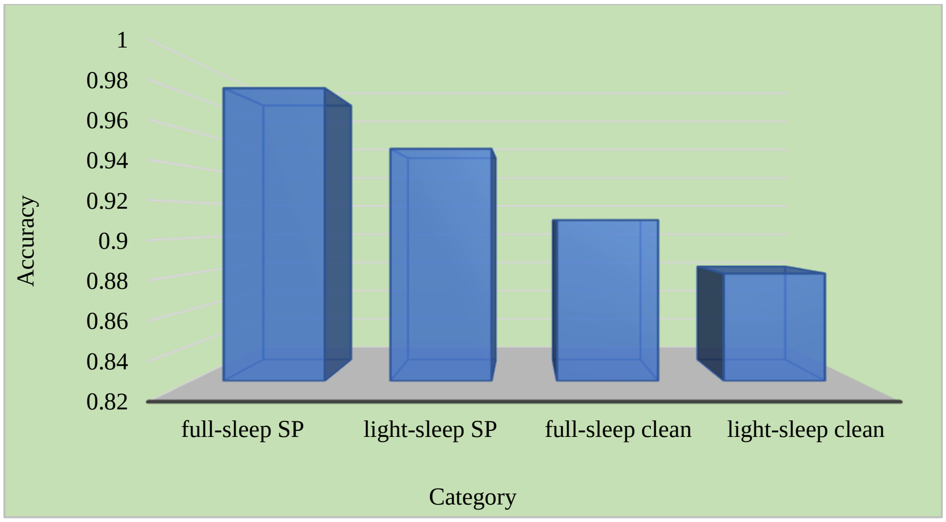


Fig. 7 Accuracy of the EBDT classifier in different scenarios

A detailed evaluation of the results for two-class superposed signals used for model classification, as presented in Table 2, indicates that EBDT scores slightly better than the other two methods, except for the AUC score. The AUC scores may differ across the ensemble models, as all ensemble models used decision trees as the base estimator, and only the learning pattern varies across ensembles.

The performance of classification models depends on the quality of the training data. The insomnia detection model using two-class signals can also be built without superposition, but at the expense of reduced model performance. Comparing the results in Tables 2 and 3, it is clear that a full-sleep stage and a light-sleep stage yield a 6%-7% increase in classification accuracy when using mixed sub-bands, rather than combining the features extracted from the individual sub-bands that constitute the sleep stages. Similarly, when examining the remaining metrics-sensitivity, specificity, precision, and F1-score-the performance increases by 3%-7%, 2%-8%, 2%-7%, and 5%-7%, respectively. In the proposed model, the evaluation parameters reveal that the full-sleep stage outperforms the light-sleep stage.

3.2. Discussions

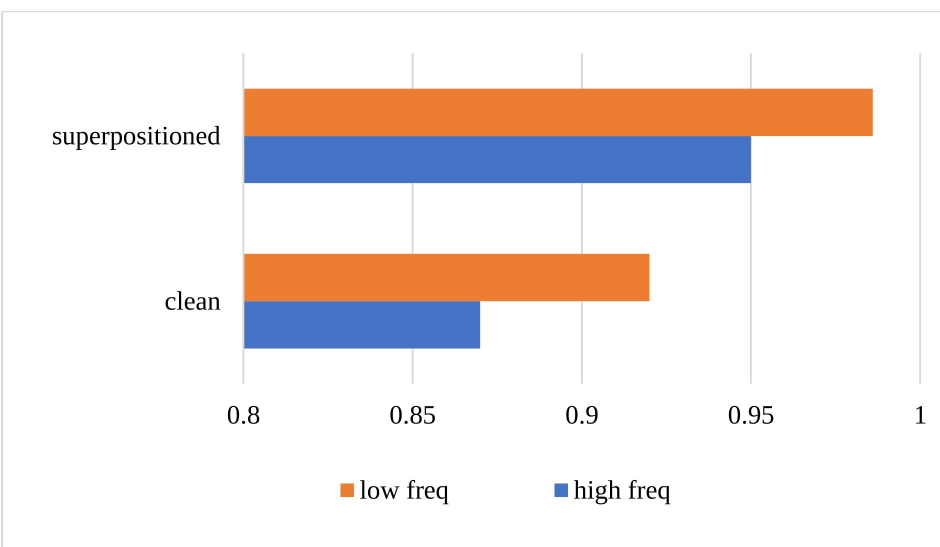
The model takes multiple composite signals from various channels and transforms them into single-frequency non-composite sub-bands, yielding decomposed versions of the raw signals. The study presents a mixed two-class approach that generates new composite signals by combining reformulated signals from individual sub-bands of the multichannel EEG, thereby reducing the complexity arising from multiple input sources. The newly generated sub-bands were subjected to time-domain segmentation to support time-frequency analysis for insomnia detection. It extracts meaningful features from the freshly segmented two-class signals and feeds them into the classifiers. During the sleep cycle, a few EEG frequencies correspond to brain activities that elicit similar responses in all subjects, irrespective of the presence of a sleep abnormality. Hence, instead of using a combined model with five frequencies to enforce individual feature extraction of segmented sub-bands, thereby increasing computation time, it is preferable to use a two-class model that combines two EEG frequencies (that shows visible characteristic variation when affected with Insomnia) to make a single signal with the characteristics of participating sub-bands, which is found to be superior in model performance and computationally more advantageous.

Hence, the proposed model follows superposition, enabling the custom combination of only significant EEG frequencies when building the detection model, thereby reducing the complexity of a traditional five-class composite model by using a two-class model for a multichannel time-frequency analysis. The classifier performance of the individual and combined sub-bands, without superposition, is detailed in Table 4.

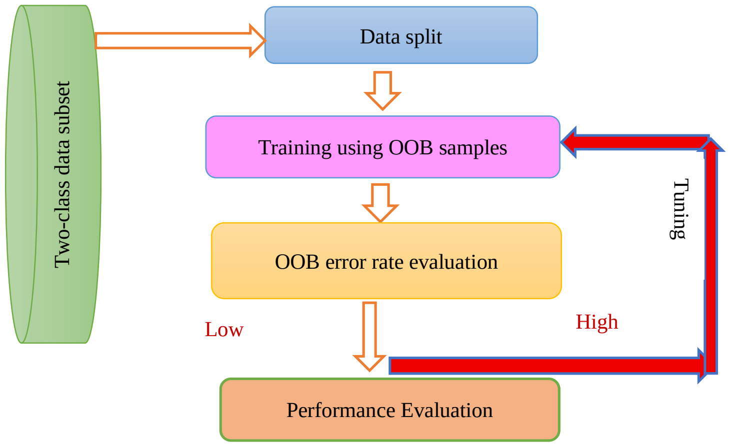
Table 4. Classifier performance for non-composite single-class sub-band signals

Single-class models	Accuracy			Sensitivity			Specificity			Precision		
	EBDT	**RF**	**GB**	**EBDT**	**RF**	**GB**	**EBDT**	**RF**	**GB**	**EBDT**	**RF**	**GB**
SB1	0.96	0.923	0.94	1.0	0.964	1.0	0.929	0.888	0.88	0.925	0.882	0.88
SB2	0.975	0.976	0.969	1.0	0.979	1.0	0.95	0.974	0.94	0.95	0.97	0.938
SB3	0.945	0.958	0.956	0.964	0.913	0.994	0.929	0.997	0.924	0.922	0.996	0.919
SB4	0.955	0.923	0.902	0.99	1.0	1.0	0.922	0.857	0.818	0.917	0.858	0.827
SB5	0.908	0.943	0.931	0.91	0.958	0.952	0.90	0.929	0.914	0.889	0.922	0.908
sub-band signals (combined)	0.828	0.846	0.847	0.853	0.912	0.936	0.803	0.78	0.758	0.812	0.806	0.794

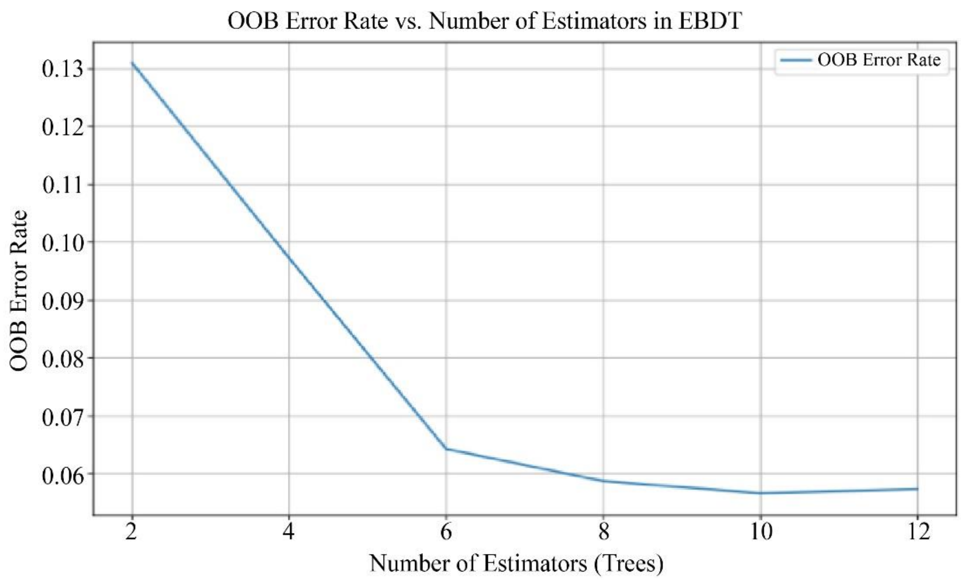
From Table 4, the classifiers respond slightly differently across Sub-Bands (SB), and this diversity is utilized in the work presented in this paper. The sub-band with the lowest classification accuracy was excluded from the two-class model creation process, and the two individual sub-bands with the closest accuracies were combined to identify the optimal two-frequency combination for the insomnia detection model. The sub-band with the lowest performance degrades the combined five-class model, as observed from the results of combined sub-band signals that include all individual sub-bands. Hence, it is convenient to use sub-bands with high diagnostic significance and better predictive performance, and to avoid sub-bands with low importance to achieve superior results. In the FS-EL model, SB1 and SB2 were used to create composite signals, while in the LS-EL model, SB3 and SB4 were used. Among the ensemble models used in the study, EBDT showed better performance and, hence, more investigation has been performed on it. While training, EBDT is tuned using Out-Of-Bag (OOB) estimates, which provide estimates for unseen samples without requiring a separate validation dataset. Generally, EBDT is built from multiple decision trees using bootstrap data points, leaving a few points aside; these left-out points are the OOB samples for that specific tree [50]. The training data for composite full-sleep, composite light-sleep, clean full-sleep, and clean light-sleep stages have been classified based on OOB scores for comparison, and the superposed data exhibit better OOB scores than the clean data, as shown in Figure 8(a). The workflow for FS-EL classification using OOB samples addressing the overfitting is depicted in Figure 8(b). The higher the OOB score, the better the classification performance [51], and it can be used as a hyperparameter when defining the model. The controlled model yields an optimized classification result, further demonstrating that the presented approach can be used to build a real-time model based on a simple, efficient, multichannel EEG architecture for precise prediction with less computation and, hence, reduced training time. A plot of OOB error rate versus number of trees, as shown in Figures 8(c) and 8(d), where the lowest error rate is visible for the full-sleep data subset with an increasing number of trees. Both two-class approaches in the model yield approximately similar results across all metrics; however, the full-sleep approach shows promising results among the newly generated two sleep stages. Furthermore, the EBDT model's performance has been re-evaluated using K-fold and stratified K-fold validation, as depicted in Figure 8(e) for K=5 in both strategies. The model's performance has been evaluated on single-class, two-class, and five-class data in both scenarios: superimposed composite signals and clean non-composite signals.



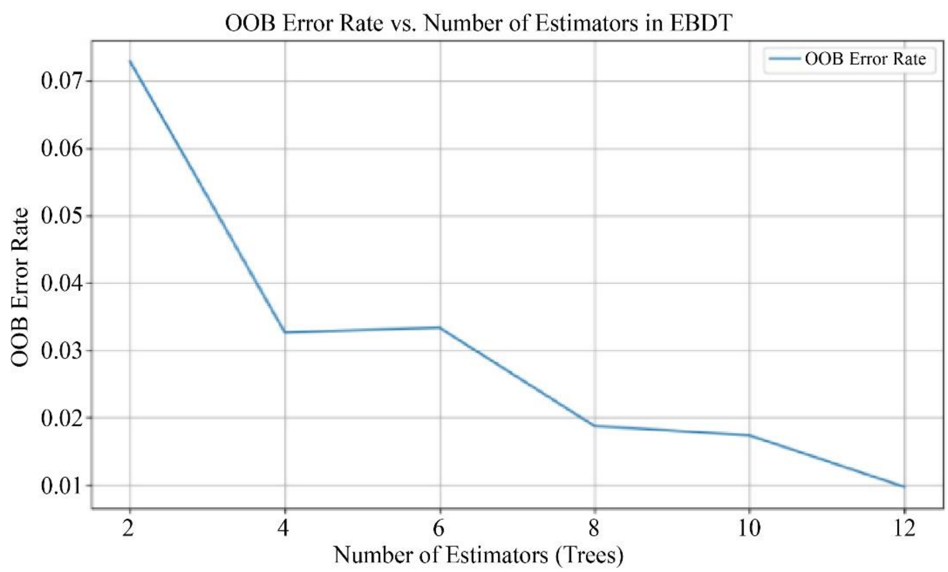
(a)


Two-class data subset
Data split
Training using OOB samples
OOB error rate evaluation
Tuning
Low
High
Performance Evaluation

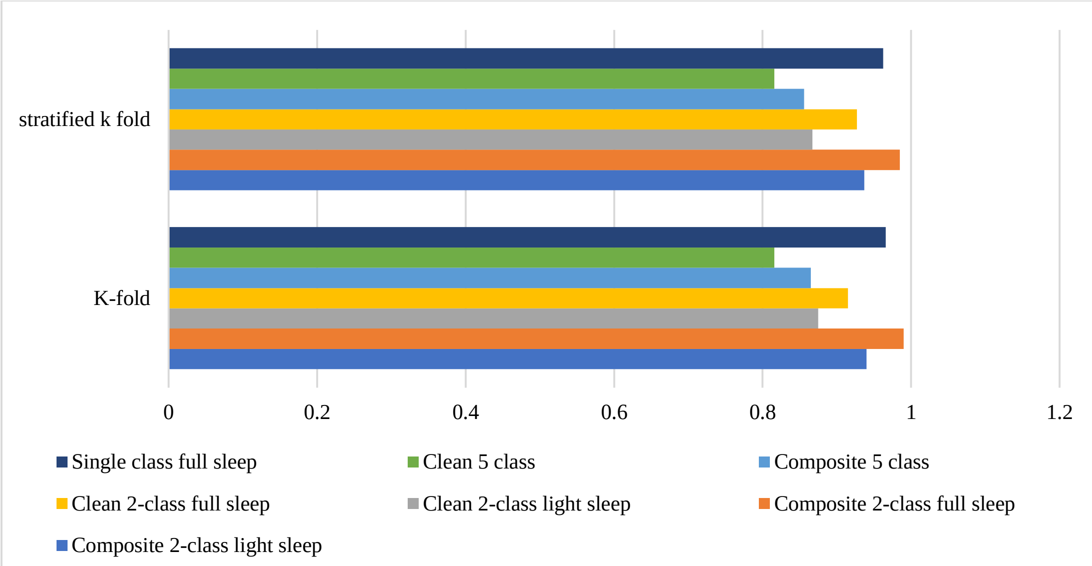
(b)


OOB Error Rate vs. Number of Estimators in EBDT
OOB Error Rate
OOB Error Rate
0.13
0.12
0.11
0.10
0.09
0.08
0.07
0.06
2
4
6
8
10
12
Number of Estimators (Trees)

(c)


OOB Error Rate vs. Number of Estimators in EBDT
OOB Error Rate
OOB Error Rate
0.07
0.06
0.05
0.04
0.03
0.02
0.01
2
4
6
8
10
12
Number of Estimators (Trees)

(d)



(e)

Fig. 8 (a) EBDT OOB score for composite and clean signals, (b) Workflow of the Fs-El classification, plot of OOB error rate Vs Number of estimators in EBDT for, (c) LS-EL, (d) FS-EL, and (e) Cross-validation on data subsets.

Table 5. (a) Comparison of existing insomnia detection models with proposed work

References	Input	Classifier	Cross-validation	Accuracy (%)
[23]	EEG	EBDT	K-fold, LOSO-CV	95.60
[29]	ECG	KNN	K-fold	97.87
[24]	EEG	EBT	K-fold	94
[25]	EEG	LSTM	-	90.9
[32]	EOG	SVM	50%-50% (training-testing) random 100 times	89.31
Proposed Work	EEG	EBDT	K-fold	99

Table 5. (b) Comparison of multichannel sleep disorder detection models with proposed work

References	Sleep disorder	Input/Number of channels	Classifier/cross-validation	Accuracy (%)
[20]	Sleep apnea	EEG/19	AdaBoost	88.3
[19]	Sleep disorder types	EEG/2	EBT/K-fold	96.63
[52]	Sleep apnea	EEG/3	SVM	99.33
[22]	Seizure	EEG/22	CNN	90
Proposed Work	Insomnia	EEG/5	EBDT/K-fold	99

From the comparative models described in Table 5(a), it is evident that most of the insomnia detection models were developed using single-channel physiological signals. The information present in other channels is not being utilized. As a result, a data deficit may arise, leading to less precise predictions in clinical diagnostic settings.

From the comparative study, a research gap has been identified in the areas of multichannel and temporal-sub-band analysis for the development of insomnia detection models, which is addressed in the work presented in this paper, achieving a greater accuracy of 0.99 with FS-EL and K-fold cross-validation. Multichannel detection models for various other sleep disorders were compared with the proposed work, as shown in Table 5(b), and it was revealed that the presented model uses only two sub-band frequencies from five channels to achieve good classification accuracy.

3.3. Advantages, Limitations, and Future Enhancements

- Employs the source of multichannel EEG data frequencies in the time domain, enabling a vast area of analytic approaches in developing the model. It also provides an analysis of smaller time segments in different EEG frequencies, satisfying the requirements for classifying datasets having a limited number of observations.
- Uses only two EEG frequencies to develop the classification model that shows reduced complexity in computation using multichannel information.
- The presented model is not dependent on sleep stage annotations, avoids human error and use of supplementary information, and hence, it has balanced and uniform training data subsets.
- Since the model has used the OOB score for hyper-tuning, it makes the model build on a controlled environment,

which makes it more generalizable and allows for efficient optimization of its performance.

- Apart from all these advantages, the general limitation of the proposed approach is the time consumption when performing manual extraction of meaningful features for generating training data subsets.
- As a future enhancement, the study can be implemented using deep learning classifiers efficiently.

4. Conclusions

This paper explores the implications of multichannel EEG for developing classification models on datasets by selectively combining sub-band frequencies. The proposed model addresses the research gap and conducts experiments to create a simple, effective model that yields more promising results than existing models, with strong coverage of data from other channels. It also produces results superior to other multichannel sleep disorder detection models. Furthermore, conventional sleep staging analysis was absent in the proposed approach, which instead examines time-frequency analysis of multichannel data to develop data subsets. The developed model was compared with composite and non-composite EEG sub-band data subsets. The excellent results were achieved with composite two-class ensemble learning models. The FS-EL EBDT classifier with five-fold cross-validation overcomes data imbalance encountered in sleep stage analysis when performing classification, achieving accuracies and AUCs of 0.99 using existing sub-band frequencies to create new composite signals comprising only prominent frequency components, thereby avoiding complex network models that generate artificial data. Detecting Insomnia using a multichannel EEG approach is carried out using other supervised ensemble learning models for all subsets discussed in the work. The results indicate that a low-frequency FS-EL data subset excelled across all ensemble models with accuracy ranging from 0.93 to 0.99, sensitivity from 0.96 to 1.0, specificity from 0.90 to 0.97, precision from 0.90 to 0.968, and F1 score from 0.93 to 0.983.

Abbreviations

EBDT/EBT- Ensemble bagged decision tree.
KNN- K nearest neighbors.
TNN- Tri-layered neural network.
LSTM- Long short-term memory.
RF-Random Forest.
SVM-Support Vector Machine.
CWT- Continuous wavelet transform.

Author contributions

SPM: Conceptualization, Methodology, design analysis, manuscript writing, and development.

TV: Supervision, manuscript development. All authors have read and accepted the manuscript.

Data Availability Statement

The data used for the work are available online on the PhysioNet website at the following URL/DOI: https://physionet.org/content/capslpdb/1.0.0/; https://doi.org/10.13026/C2VC79.

References

[1] Harvey R. Colten, and Bruce M. Altevogt, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*, Institute of Medicine (US) Committee on Sleep Medicine and Research, pp. 55-136, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[2] Milena K. Pavlova, and Véronique Latreille, "Sleep Disorders," *The American Journal of Medicine*, vol. 132, no. 3, pp. 292-299, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Michael J. Thorpy, "Classification of Sleep Disorders," *Neurotherapeutics*, vol. 9, no. 4, pp. 687-701, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[4] Thomas Roth, "Insomnia: Definition, Prevalence, Etiology, and Consequences," *Journal of Clinical Sleep Medicine*, vol. 3, no. 5 suppl, pp. S7-S10, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[5] Michael J. Sateia et al., "Evaluation of Chronic Insomnia. An American Academy of Sleep Medicine Review," *Sleep*, vol. 23, no. 2, pp. 243-308, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jessica Vensel Rundo, and Ralph Downey III, *Polysomnography*, Handbook of Clinical Neurology, vol. 160, pp. 381-92, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] Wu Wen, "Sleep Quality Detection based on EEG Signals using Transfer Support Vector Machine Algorithm," *Frontiers in Neuroscience*, vol. 15, 2021. [Google Scholar]

[8] Yi-Hsuan Cheng, Margaret Lech, and Richardt Howard Wilkinson, "Simultaneous Sleep Stage and Sleep Disorder Detection from Multimodal Sensors using Deep Learning," *Sensors*, vol. 23, no. 7, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Shuting Xu et al., "A Review of Automated Sleep Disorder Detection," *Computers in Biology and Medicine*, vol. 150, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Erdenebayar Urtnasan, Eun Yeon Joo, and Kyu Hee Lee, "AI-Enabled Algorithm for Automatic Classification of Sleep Disorders based on Single-Lead Electrocardiogram," *Diagnostics*, vol. 11, no. 11, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Michael T. Smith et al., "Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Systematic Review, Meta-Analysis, and Grade Assessment," *Journal of Clinical Sleep Medicine*, vol. 14, no. 7, pp. 1209-1230, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[12] Michael T. Smith et al., "0328 Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Clinical Practice Guideline," *Sleep*, vol. 42, no. 1, pp. A134-A135, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[13] Desta Fekedulegn et al., "Actigraphy-Based Assessment of Sleep Parameters," *Annals of Work Exposures and Health*, vol. 64, no. 4, pp. 350-367, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ian G. Campbell, "EEG Recording and Analysis for Sleep Research," *Current Protocols in Mouse Biology*, vol. 49, no. 1, pp. 1-21, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[15] Fabien Lotte, "A Tutorial on EEG Signal-Processing Techniques for Mental-State Recognition in Brain-Computer Interfaces," *Guide to Brain-Computer Music Interfacing*, pp. 133-161, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[16] Fahd A. Alturki et al., "EEG Signal Analysis for Diagnosing Neurological Disorders using Discrete Wavelet Transform and Intelligent Techniques," *Sensors*, vol. 20, no. 9, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] D. Puthankattil Subha et al., "EEG Signal Analysis: A Survey," *Journal of Medical Systems*, vol. 34, no. 2, pp. 195-212, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[18] Ramina Behzad, and Aida Behzad, "The Role of EEG in the Diagnosis and Management of Patients with Sleep Disorders," *Journal of Behavioral and Brain Science*, vol. 11, no.10, pp. 257-266, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Manish Sharma et al., "Automated Identification of Sleep Disorder Types using Triplet Half-Band Filter and Ensemble Machine Learning Techniques with Eeg Signals," *Electronics*, vol. 10, no. 13, pp. 1-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Dongyeop Kim et al., "Machine-Learning-Based Classification of Obstructive Sleep Apnea using 19-Channel Sleep EEG Data," *Sleep Medicine*, vol. 124, pp. 323-330, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Santosh Kumar Satapathy et al., "Machine Learning-Empowered Sleep Staging Classification using Multi-Modality Signals," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, pp. 1-29, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Xiaoyan Wei et al., "Automatic Seizure Detection using Three-Dimensional CNN based on Multi-Channel EEG," *BMC Medical Informatics and Decision Making*, vol. 18, no. S5, pp. 71-80, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] Manish Sharma, Virendra Patel, and U. Rajendra Acharya, "Automated Identification of Insomnia using Optimal Bi-Orthogonal Wavelet Transform Technique with Single-Channel EEG Signals," *Knowledge-Based Systems*, vol. 224, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Manish Sharma et al., "Automated Insomnia Detection using Wavelet Scattering Network Technique with Single-Channel EEG Signals," *Engineering Applications of Artificial Intelligence*, vol**.** 126, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Wei Qu et al., "Single-Channel EEG based Insomnia Detection with Domain Adaptation," *Computers in Biology and Medicine*, vol. 139, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Yang B, Liu H, "Automatic Identification of Insomnia based on Single-Channel EEG Labelled with Sleep Stage Annotations," *IEEE Access*, vol. 8, pp. 104281-104291, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Shivam Tiwari, Deepak Arora, and Vishal Nagar, "Detection of Insomnia using Advanced Complexity and Entropy Features of Sleep Stage Data of EEG Recordings," *Measurement: Sensors*, vol. 24, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Mohd Maroof Siddiqui, Geetika Srivastava, and Syed Hasan Saeed, "Diagnosis of Insomnia Sleep Disorder using Short Time Frequency Analysis of PSD Approach Applied on EEG Signal using Channel ROC-LOC," *Sleep Science*, vol. 9, no. 3, pp. 186-91, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[29] Manish Sharma, Harsh S. Dhiman, and U. Rajendra Acharya, "Automatic Identification of Insomnia using Optimal Antisymmetric Biorthogonal Wavelet Filter Bank with ECG Signals," *Computers in Biology and Medicine*, vol. 131, pp. 1-33, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[30] Kamlesh Kumar et al., "INSOMNet: Automated Insomnia Detection using Scalogram and Deep Neural Networks with ECG Signals," *Medical Engineering & Physics*, vol. 119, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] Nishant Sharma et al., "Automated Accurate Insomnia Detection System using Wavelet Scattering Method using ECG Signals," *Applied Intelligence*, vol. 54, no. 4, pp. 3464-3481, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Chih-En Kuo, and Guan-Ting Chen, "A Short-Time Insomnia Detection System based on Sleep EOG with RCMSE Analysis," *IEEE Access*, vol. 8, pp. 69763-36773, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[33] Shitanshu Kusmakar et al., "A Machine Learning Model for Multi-Night Actigraphic Detection of Chronic Insomnia: Development and Validation of a Pre-Screening Tool," *Royal Society Open Science*, vol. 8, no. 6, pp. 1-17, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[34] Maia Angelova et al., "Automated Method for Detecting Acute Insomnia using Multi-Night Actigraphy Data," *IEEE Access*, vol. 8, pp. 74413-74422, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[35] S. Rani et al., "Differentiating Acute from Chronic Insomnia with Machine Learning from Actigraphy Time Series Data," *Frontiers in Network Physiology*, vol. 2, 2022. [Google Scholar]

[36] Mahan Choudhury et al., "Explainable AI-Driven Scalogram Analysis and Optimized Transfer Learning for Sleep Apnea Detection with Single-Lead Electrocardiograms," *Computers in Biology and Medicine*, vol. 187, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[37] Duc Thien Pham, and Roman Mouček, "Efficient Sleep Apnea Detection using Single-Lead ECG: A CNN-Transformer-LSTM Approach," *Computers in Biology and Medicine*, vol. 196, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[38] Md Belal Bin Heyat et al., "Sleep Bruxism Detection using Decision Tree Method by the Combination of C4-P4 and C4-A1 Channels of Scalp EEG," *IEEE Access*, vol. 7, pp. 102542-102553, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[39] Serap Aydın, Hamdi Melih Saraoğlu, and Sadık Kara, "Singular Spectrum Analysis of Sleep EEG in Insomnia," *Journal of Medical Systems*, vol. 35, no. 4, pp. 457-461, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[40] Ary L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[41] Mario Giovanni Terzano et al., "Atlas, Rules, and Recording Techniques for the Scoring of Cyclic Alternating Pattern (CAP) in Human Sleep," *Sleep Medicine*, vol. 2, no. 6, pp. 537-553, 2001. [CrossRef] [Google Scholar] [Publisher Link]

[42] Hyunsoo Yu et al., "Deep Neural Network-Based Empirical Mode Decomposition for Motor Imagery EEG Classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 3647-3656, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[43] Jasmin Kevric, and Abdulhamit Subasi, "Comparison of Signal Decomposition Methods in Classification of EEG Signals for Motor-Imagery BCI System," *Biomedical Signal Processing and Control*, vol. 31, pp. 398-406, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[44] Ayşe Nur Camgöz et al., "Analysis of Electroencephalography (EEG) Signals and Adapting to Systems Theory Principles," *Journal of Complexity in Health Sciences*, vol. 4, no. 2, pp. 39-44, 2021. [CrossRef] [Google Scholar] [publisher link]

[45] Sotiris B Kotsiantis, George E. Tsekouras, and Panayiotis E. Pintelas, "Bagging Model Trees for Classification Problems," *Panhellenic Conference on Informatics*, pp. 328-337, 2005. [CrossRef] [Google Scholar] [publisher link]

[46] Robi Polikar, "Ensemble based Systems in Decision Making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21-45, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[47] Gérard Biau, "Analysis of a Random Forests Model," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 1063-1095, 2012. [Google Scholar] [Publisher Link]

[48] Hasan Ahmed Salman, Ali Kalakech, and Amani Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69-79, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[49] Maria D. Guillen, Juan Aparicio, and Miriam Esteve, "Gradient Tree Boosting and the Estimation of Production Frontiers," *Expert Systems with Applications*, vol. 214, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[50] Helen L. Smith et al., "Out of (the) Bag-Encoding Categorical Predictors Impacts Out-of-Bag Samples," *PeerJ Computer Science*, vol. 10, pp. 1-18, 2024.[CrossRef] [Google Scholar] [Publisher Link]

[51] Yongchan Kwon, and James Zou, "Data-OOB: Out-of-Bag Estimate as a Simple and Efficient Data Value," *Proceedings of the 40th International Conference on Machine Learning*, pp. 18135-18152, 2023. [Google Scholar] [Publisher Link]

[52] Behnam Gholami et al., "Sleep Apnea Diagnosis using Complexity Features of EEG Signals," *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pp. 74-83, 2022. [CrossRef] [Google Scholar] [Publisher Link]