DOI https://doi.org/10.24297/ijct.v25i.9809

Detection of Deep Sleep Stages in Multi-Channel EEG Signals Based on Spectral Feature Quantification Methods

Jinsha Liu[1], *Chengyuan Shen*[3], *Liangyu Zhao*[4] ,*Taiyo Maeda*[1] *and Jianting Cao* [1,2] *

[1]Graduate School of Engineering, Saitama Institute of Technology, Fukaya City, Saitama, Japan

[2]RIKEN Center for Advanced Intelligence Project (AIP), Chuo-ku, Tokyo, Japan

[3]Keio University, Yokohama City, Kanagawa, Japan

[4]More & More Company Limited, Zushi City, 2490006, Kanagawa, Japan

Email addresses ı4001bon@sit.ac.jp, chengyuanshen@keio.jp, zhao@momore.co.jp, maedat@sit.ac.jp, cao@sit.ac.jp

Abstract

Deep sleep is essential for physical recovery and cognitive function. Accurate detection is crucial for evaluating sleep quality and diagnosing related disorders. Electroencephalography (EEG) remains the most reliable tool for assessing deep sleep. However, while deep learning methods have shown high performance, they often suffer from limited interpretability and require substantial computational resources, which can hinder real-time clinical applications. This study proposes a novel multi-channel EEG analysis approach combining Turning Tangent Empirical Mode Decomposition (2T-EMD) and Multi-Taper Power Spectral Density (MT-PSD) to extract physiologically meaningful spectral features, followed by classification using a Random Forest (RF) model. The method was validated on 61 recordings from 41 participants, achieving 95.57% accuracy, 97.35% recall, 97.58% F1-score, and 99.00% AUC. Compared with existing approaches, our method demonstrates (1) physiologically interpretable feature extraction; (2) favorable computational efficiency under our experimental setup; and (3) robust generalizability across different demographics, indicating its potential utility in clinical sleep monitoring scenarios.

Keywords: EEG, BCI, Sleep Monitoring, Deep learning, Data Processing

I.Introduction

Deep sleep is a critical phase for human recovery and restoration, playing a vital role in maintaining both physical health and cognitive function. Accurate detection of the deep sleep stage is crucial for evaluating sleep quality and diagnosing related disordersChokroverty, 2010; Fattinger et al., 2017. Among various methods for monitoring sleep states, Electroencephalogram (EEG) signals are widely used for sleep stage classification Aboalayon et al., 2016; Eldele et al., 2021, as they directly reflect the brain's activity. EEG is also employed in the diagnosis and research of neurological disorders such as epilepsy Andrzejak et al., 2012; Zhao et al., 2022 and brain death diseaseCao and Chen, 2008; Chen et al., 2019. The distinctive slow-wave activity in the EEG during deep sleep provides reliable physiological evidence for identifying this stage.

In recent years, with the development of artificial intelligence technologies, significant progress has been made in sleep stage classification using deep learning methods. Researchers have proposed various deep learning technology Gabriel-Alexandru and Mihaela-Ruxandra, 2024; Lozzi et al., 2024 and deep learning frameworks integrating attention mechanisms Jin and Jia, 2023; Wu et al., 2024, which have demonstrated excellent performance in sleep stage classification tasks. Additionally, methods for feature extraction based on multi-channel EEG signals and the application of machine learning algorithms in sleep stage classification have been exploredChambon et al., 2018; Phan et al., 2021; Subasi and Gursoy, 2010. However, these methods face two main challenges in practical clinical applications: first, real-time deep sleep detection requires fast computational capabilities, but complex deep learning models often demand substantial computational resources; second, deep learning methods typically extract high-dimensional features for classification, but these features lack clear physiological explanations, limiting their applicability in medical fields that require rigorous theoretical foundations. To address these issues, this paper proposes a novel method for deep sleep stage detection using multi-channel EEG signals based on spectral feature quantification. The method integrates a Turning Tangent Empirical Mode Decomposition (2T-EMD) algorithm Fleureau et al., 2010, Multitaper Power Spectral Density (MT-PSD) technique Babadi and Brown, 2014, and a Random Forest classifier to achieve accurate identification of the deep sleep stage. Validation of this method on EEG data from 41 participants showed excellent recognition accuracy in the deep sleep detection task. Compared with existing approaches, the proposed method offers several advantages, including strong interpretability, high generalizability, and model stability, while also maintaining high computational efficiency, making it suitable for real-time deployment in clinical settings.

The main contributions of this work are summarized as follows:

1. We propose a novel deep sleep stage detection method that combines 2T-EMD, MT-PSD, and Random Forest, enabling effective spectral feature quantification from multi-channel EEG signals.
2. We validate the proposed method on EEG recordings from 41 participants, achieving excellent recognition accuracy in deep sleep detection.
3. The proposed approach demonstrates strong interpretability, high generalizability, and stable performance across participants.
4. The method maintains high computational efficiency, making it suitable for real-time applications in clinical settings.

The remainder of this paper is organized as follows: Section introduces the proposed method in detail, including the 2T-EMD algorithm, MT-PSD analysis technique, and Random Forest classifier. Section presents the experimental design and result analysis. Section provides an in-depth discussion of the advantages and limitations of the proposed method. Finally, Section concludes the paper and discusses potential future research directions.

II.Methods

a. Turning Tangent Empirical Mode Decomposition

EEG signals are inherently non-stationary, with their amplitude typically varying over time. Traditional signal processing techniques, such as the Fast Fourier Transform (FFT) Heckbert, 1995, are often unable to accurately represent the frequency spectrum of such signals due to the assumption of signal stationarity. Given this challenge, methods such as Empirical Mode Decomposition (EMD) Flandrin et al., 2004 are commonly employed as preprocessing steps to better handle non-stationary signals like EEG.

EMD is a classical algorithm for decomposing non-stationary signals into a series of Intrinsic Mode Functions (IMFs) that capture frequency-specific components, enabling multi-scale feature extraction. These IMFs are arranged in order of decreasing frequency content, with the final component typically representing the residual or trend of the signal. After the EMD process, the original non-stationary signal is transformed into a series of pseudo-stationary signals, enabling more stable frequency characteristics that are suitable for further processing, such as FFT analysis.

However, traditional EMD has significant limitations when applied to multi-channel signals, such as mode mixing and high computational complexity. Common variants of EMD, such as Ensemble EMD (EEMD) and Complete EEMD with Adaptive Noise (CEEMDAN), mitigate mode mixing by introducing noise-assisted analysis. Nevertheless, these methods require numerous ensemble averages, resulting in prohibitively high computational costs that cannot meet the core requirement of real-time processing for multi-channel EEG in this studyWang et al., 2014. Furthermore, since EEMD and CEEMDAN decompose each channel independently, they fail to address the critical issue of mode misalignment across channels.

To address these issues simultaneously, we utilized an improved Turning Tangent Empirical Mode Decomposition (2T-EMD) algorithmFleureau et al., 2010. The primary reasons for selecting 2T-EMD over other variants are its distinct advantages: First, its algorithm incorporates a real-time calculation mechanism that dynamically tracks the time-domain energy variations of the signal, resulting in computational efficiency significantly higher than that of EEMD/CEEMDAN, making it highly suitable for the rapid processing of multi-channel EEG signals. Second, the decomposition mechanism of 2T-EMD better preserves the temporal consistency of corresponding IMF components across different channels, effectively alleviating mode misalignment and providing a more reliable foundation for subsequent multi-channel joint analysis.

By adopting this approach, we can achieve accurate decomposition of multi-channel EEG signals efficiently, providing a reliable and practical foundation for subsequent signal processing and analysis tasks in real-time scenarios.

b. Multitaper Power Spectral Density

The Multitaper Power Spectral Density (MT-PSD) is a highly accurate spectral estimation technique that utilizes a set of orthogonal taper functions (Slepian sequences) to analyze signals Babadi and Brown, 2014. This method effectively reduces spectral leakage and provides stable and high-resolution power spectral density (PSD) estimates. Compared to traditional FFT, MT-PSD offers superior spectral estimation performance and lower variance, making it particularly advantageous for analyzing non-stationary signals such as EEG. In this study, MT-PSD is applied to analyze the pseudo-stationary signals obtained from 2T-EMD decomposition, facilitating the extraction of critical features associated with deep sleep stages. The fundamental concept of MT-PSD involves weighting a signal with a series of orthogonal taper functions, thereby minimizing spectral leakage while maintaining robust frequency resolution. For a given time series $x(n)$, the PSD estimate using Multitaper method can be expressed asYoungworth et al., 2005:

$$\hat{S}(f) = \frac{1}{K} \sum_{k=1}^{K} |\mathcal{F}\big(x(n) \cdot w_k(n)\big)|^2 \tag{1}$$

where $w_k(n)$ represents the kth taper function, K denotes the number of taper functions used, and F indicates the Fourier transform operation. MT-PSD significantly reduces the variance of spectral estimates by averaging results obtained from multiple taper functions. Each taper function is optimally designed to minimize spectral leakage within a specified time-bandwidth product, making MT-PSD particularly effective for analyzing non-stationary time series like EEG signals.

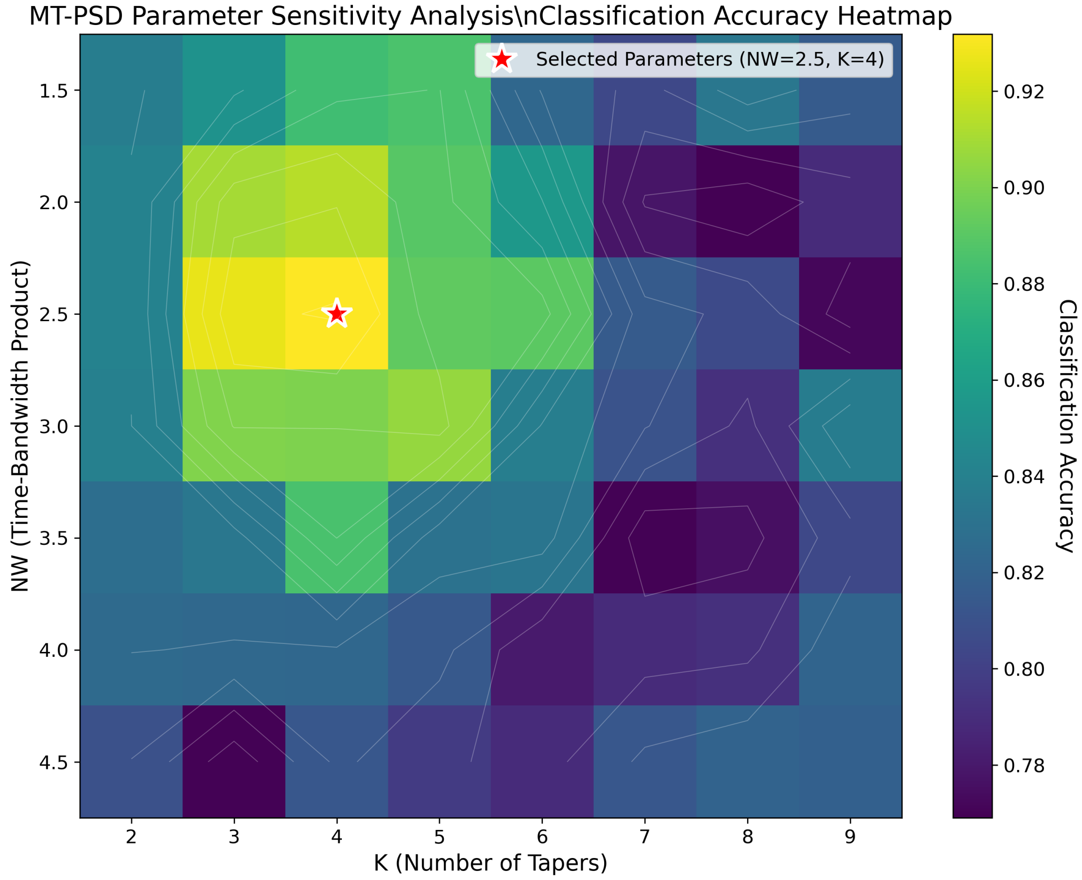


Figure 1: mtpsd parameter sensitivity

In this study, MT-PSD was applied to the IMFs obtained from the 2T-EMD decomposition for spectral analysis. Specifically, the time-bandwidth parameter $NW = 2.5$ (time-bandwidth product) and the number of taper functions $K = 4$ were selected, as these values demonstrated optimal performance during experimental validation.The PSD features extracted using Multitaper method clearly captured the low-frequency energy characteristics unique to the deep sleep stage, providing reliable input for subsequent classification tasks. Furthermore, MT-PSD's computational efficiency has been optimized for practical applications, particularly when handling high-dimensional multi-channel EEG data. The method effectively balances computational complexity and accuracy, making it well-suited for real-time processing. By integrating MT-PSDderived spectral features with a Random Forest classifierParmar et al., 2018, our approach significantly enhances the accuracy and robustness of deep sleep stage detection.

c. Random Forest

Random Forest (RF) is a powerful ensemble learning classification method that achieves high accuracy by constructing multiple decision trees and combining their predictions. It is particularly well-suited for handling high-dimensional data and noisy signals, offering strong noise resistance and excellent generalization performance. In this study, RF is employed to classify spectral features derived from the MT-PSD analysis for detecting deep sleep stages. The construction of a RF involves the following steps:

1. From a training set containing N samples, each represented by an M-dimensional feature vector, K subsets are generated via bootstrapping (random sampling with replacement). Each subset is used to train an individual decision

tree.

2. During tree construction, a random subset of features (typically $\sqrt{M}$) is selected at each split point, and split conditions are based only on this subset.

3. Gini impurity is used to evaluate the split quality, defined as:

$$G = 1 - \sum_{i=1}^{C} p_i^2 \tag{2}$$

where C is the total number of classes, and p_i represents the proportion of samples belonging to the i-th class.

The final classification result is determined by majority voting among all decision trees:

$$\hat{y} = \text{majority\_vote}\left(T_1(x), T_2(x), \ldots, T_K(x)\right) \tag{3}$$

where $T_k(x)$ is the prediction of the k-th decision tree, and $\hat{y}$ is the final predicted class.

For this study, the RF model was optimized to suit the characteristics of EEG signals. Each input sample comprises 70 spectral features obtained through MT-PSD analysis, reflecting energy distribution across various frequency bands. One of the key advantages of RF is its ability to evaluate feature importance, offering valuable interpretability.

By analyzing feature importance, we identified frequency bands most relevant to deep sleep classification, such as specific low-frequency energy features. This not only enhances the interpretability of the proposed method but also provides meaningful insights for clinical research and practical applications.

III.Experiment and Results

a. The Sleep-EDF Database Expanded

The Sleep-EDF dataset contains 197 full-night Polysomnographic (PSG) recordings, including EEG, EOG, chin EMG, and event markers Kemp et al., 2000. These recordings are extensively used for sleep stage classification and neurological disease studies. Some recordings also include respiratory flow and body temperature data. The corresponding sleep stages were manually annotated by trained technologists according to the Rechtschaffen and Kales manual. All EEG signals are recorded from multiple electrode positions (e.g., Fpz-Cz and Pz-Oz), and the dataset follows the EDF+ header specifications, with unrecorded signals removed from the ST*PSG.edf files. Each sleep stage recording includes various stages such as W, R, 1, 2, 3, 4. The majority of the dataset consists of full-night PSG recordings, obtained under various conditions, including in-home data collection.

The dataset is divided into two main parts:

1.The Sleep Cassette Study and Data

This part consists of 153 SC files (SC stands for Sleep Cassette) obtained from a study conducted between 1987 and 1991. The participants were healthy Caucasians aged between 25 and 101 years, and they were not taking any sleep-related medications. The study aimed to investigate the effects of age on sleep patterns. Data was recorded at participants' homes, with two PSG recordings conducted during both daytime and nighttime, each lasting about 20 hours. Participants wore a Walkman-type cassette recorder, engaging in normal activities while data was recorded continuously. Due to equipment failure, the first night's data of participants 36 and 52, and the second night of participant 13 were lost Kemp et al., 2000. Additionally, data for participants numbered 39, 68, 69, 78, and 79 were missing from the dataset. Ultimately, the dataset consists of PSG recordings from 77 participants. Both EEG and EOG signals were sampled at 100 Hz. Chin EMG signals were high-pass filtered, rectified, and low-pass filtered, then sampled at 1 Hz and represented in RMS (root mean square) units. Respiratory flow, rectal temperature, and event markers were also sampled at 1 Hz.

For the present study, we only used the Sleep Cassette recordings, excluding the Telemetry recordings to avoid the confounding effects of Temazepam. After applying the selection criterion that required each subject to have at least 1,500 seconds (50 epochs) of deep sleep, a total of 61 valid PSG sessions were retained. Among them, deep sleep epochs (positive samples) accounted for approximately **10.8%** of the total, while non-deep sleep epochs (negative samples) accounted for the remaining **89.2%**. This ensures sufficient representation of both classes for training and evaluation, although the imbalance reflects the natural distribution of sleep stages.

2.Sleep Telemetry Study and Data

This part consists of data from a 1994 study involving 22 healthy Caucasian male and female participants, which investigated the effects of Temazepam on sleep. No other medications were used by the participants, and the data was recorded in a hospital setting. Each participant had minor difficulties falling asleep, but was otherwise healthy. PSG data was recorded over two nights, one with Temazepam and the other with a placebo. Participants wore a miniature telemetry system, and the signal quality was excellent Kemp et al., 2000. Although these recordings are of high quality, they were excluded from this study in order to avoid drug-related bias.

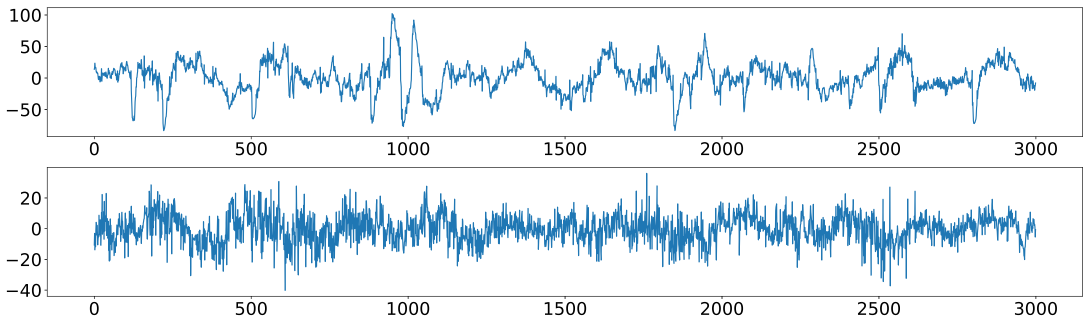


Figure 2: An example of EEG signal in sleep stage 1.

Figure. 2 provides an example of EEG signal which is in sleep stage1.

Figure. 3 presents a comparison of computational efficiency among 2T-EMD, EMD, EEMD, and CEEMDAN. As illustrated, 2T-EMD substantially reduces processing time while maintaining comparable decomposition accuracy. In contrast, EEMD and CEEMDAN, which require multiple ensemble iterations, incur significantly higher computational costs, particularly for multi-channel EEG data. The figure demonstrates that 2T-EMD scales more efficiently with the number of channels, making it more suitable for real-time and resource-constrained applications in both research and clinical settings. For this study, we focused only on the data from the "Sleep Cassette Study and Data" part of the dataset to avoid potential confounding effects of medication on EEG signals. To ensure high data quality, only recordings with a deep sleep duration of over 1500 seconds were selected for analysis. Shorter durations were excluded as they did not provide sufficient data for training and testing. In total, 61 recordings from 41 participants were selected for feature extraction, model training, and testing.

b. Experiment Flow

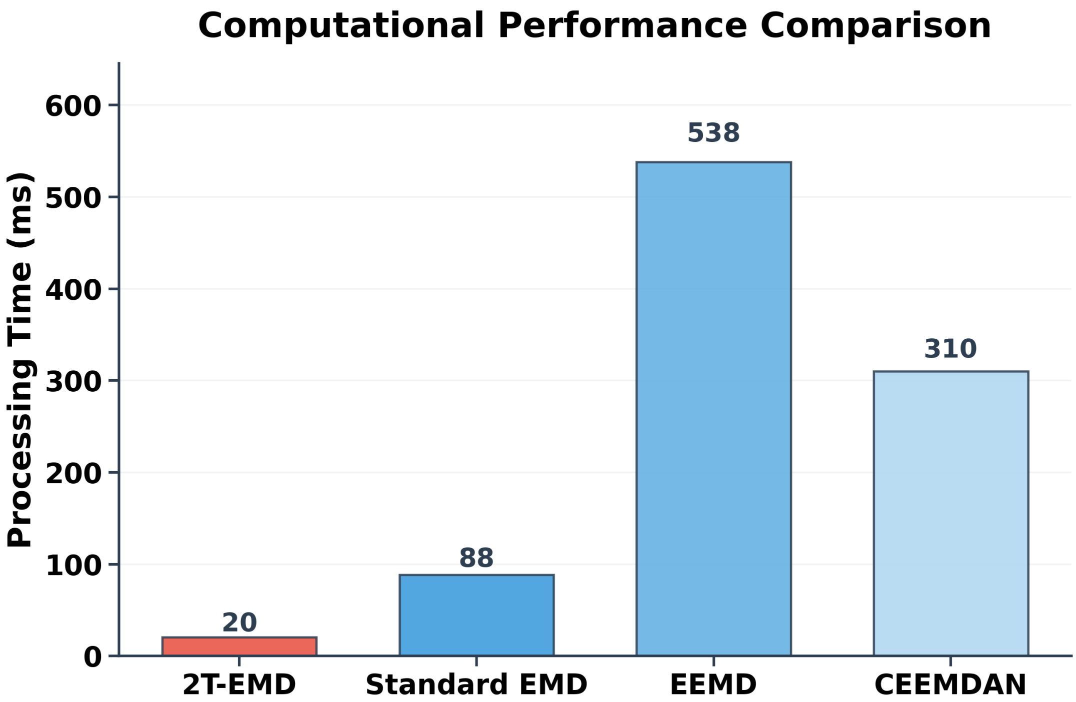
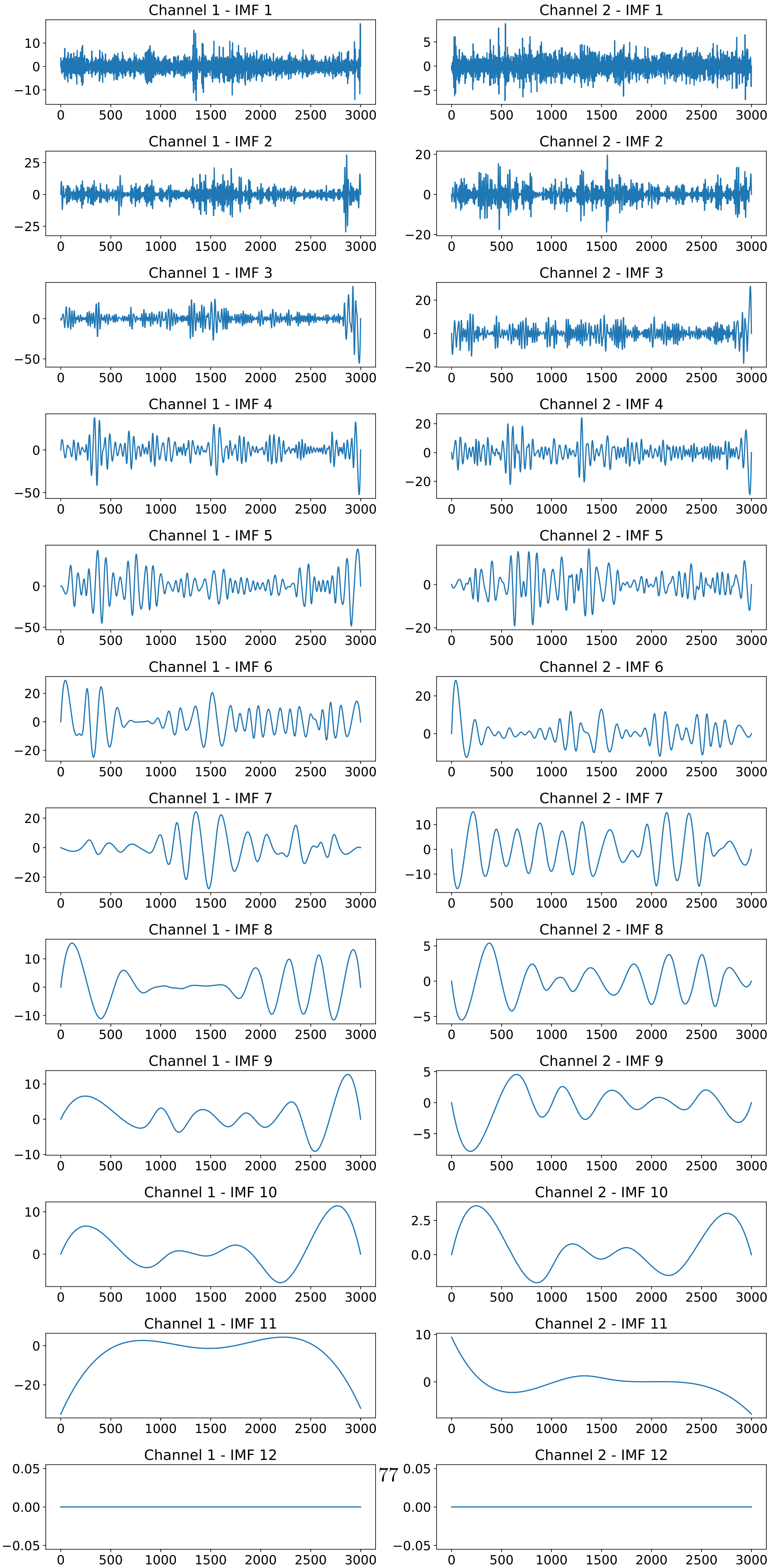


Figure 3: Computational Efficiency Comparison of 2T-EMD.

1. In this study, we performed a systematic preprocessing of the raw EEG data. Based on the manually annotated sleep stage labels provided in the dataset, we segmented the EEG data of each participant into 30-second epochs. This duration was chosen according to the standard epoch length used in traditional sleep staging protocols. The 30-second epoch not only adheres to international sleep analysis standards but also balances the time resolution of the signals with the classification task's accuracy requirements. For each

30-second epoch, we extracted the corresponding label information and classified the data according to the various sleep stages. Specifically, we focused on distinguishing between deep sleep stages (N3 and N4) and other sleep stages, laying the foundation for the subsequent binary classification task. In the classification process, deep sleep data (N3 and N4 stages) were merged into positive labels, while data from other sleep stages, including Rapid Eye Movement (REM, R stage), light sleep (N1, N2 stages), and intermediate stages, were grouped as negative labels.

2. Considering the individual differences in the duration of deep sleep, we established stringent data selection criteria to ensure the reliability and representativeness of the dataset. We first calculated the total duration of each sleep stage for each subject by determining the number of epochs corresponding to each stage. To ensure that the selected data could accurately represent deep sleep characteristics, we set the inclusion criteria: the total duration of deep sleep (Stages 3 and 4) for each subject had to be at least 1500 seconds (i.e., 50 epochs). This threshold was based on the typical duration of deep sleep and aimed to ensure that the data retained in the analysis adequately captured the key features of deep sleep stages. To eliminate potential time-sequencing effects that might influence the results, we applied a randomization procedure to the data. Specifically, we shuffled the 50 epochs of deep sleep data using a fixed random seed and then randomly selected 40 epochs for further analysis. This approach was designed to prevent the classifier from overfitting to the temporal features of the dataset, thereby improving the generalization performance of the model. After applying these quality control measures, we retained the data from subjects who met the selection criteria for subsequent feature extraction, training, and testing.


Channel 1 - IMF 1
Channel 2 - IMF 1
Channel 1 - IMF 2
Channel 2 - IMF 2
Channel 1 - IMF 3
Channel 2 - IMF 3
Channel 1 - IMF 4
Channel 2 - IMF 4
Channel 1 - IMF 5
Channel 2 - IMF 5
Channel 1 - IMF 6
Channel 2 - IMF 6
Channel 1 - IMF 7
Channel 2 - IMF 7
Channel 1 - IMF 8
Channel 2 - IMF 8
Channel 1 - IMF 9
Channel 2 - IMF 9
Channel 1 - IMF 10
Channel 2 - IMF 10
Channel 1 - IMF 11
Channel 2 - IMF 11
Channel 1 - IMF 12
Channel 2 - IMF 12

3. During the feature extraction phase, we adopted a systematic data processing pipeline to ensure that the extracted features accurately reflect the characteristics of deep sleep EEG signals. For each subject's data that met the inclusion criteria, we first applied a randomization procedure within each sleep stage using a fixed random seed to eliminate potential time-sequencing effects on the results. We then sequentially selected 5 epochs of data at a time for processing. The specific steps are as follows:

1). 2T-EMD Decomposition: We applied the 2T-EMD method to decompose the EEG data from two channels into a set of intrinsic mode functions (IMFs). The 2T-EMD method effectively handles the non-stationary nature of EEG signals by decomposing them into frequency components, providing a clearer representation of the signal. Figure. 4 shows an example of 2 channel EEG signals decomposed by 2T-EMD.

2). MT-PSD Analysis: For each IMF, we used a 2-second window with 1-second overlap to segment the signal. Then, we applied the MT-PSD to compute the power spectral density for each segment. MT-PSD uses a set of orthogonal tapers to analyze the signal, which effectively reduces spectral leakage and provides high-precision spectral estimates.

3). Spectral Density Aggregation: For each segment, the computed spectral density is aggregated across the same time window, yielding the overall spectral characteristics of that segment. Each segment is 2 seconds long with a sampling frequency of 100Hz, resulting in 200 data points per segment. The frequency resolution after MT-PSD is 0.5Hz, and since the effective frequency range of EEG signals is 0.5-35Hz, we select the first 70 frequency points within the 0.5-35Hz range.

4). Averaging Spectral Features: All spectral density except the last one in the same time segment are summed together. The spectral features from all the time segments of the entire signal are summed and averaged to obtain a representative energy feature for each 30-second window. This process helps reduce noise interference and improves the stability and reliability of the features.

5). Dataset Construction: The aforementioned processing steps were applied to all subjects' data, resulting in a dataset containing spectral features. This dataset includes 4 sleep stages, 61 pieces, 40 epochs per piece, data from 2 EEG channels, and 70 frequency features for each epoch. In this study, the frequency range of 0.5–70 Hz was selected for spectral feature extraction. This range covers all major EEG bands relevant to sleep analysis, including delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and low gamma (30–70 Hz), thereby capturing both slow-wave activity characteristic of deep sleep and higher-frequency components associated with wakefulness or REM sleep. Frequencies below 0.5 Hz were excluded to reduce the influence of baseline drift, respiratory and cardiac artifacts, while frequencies above 70 Hz were omitted to minimize contamination from muscle activity and environmental noise. This selection ensures retention of physiologically meaningful signals while maintaining robustness against common EEG artifacts, and aligns with prior studies in EEG spectral analysis Babadi and Brown, 2014; Subasi and Gursoy, 2010.This dataset was used for subsequent classifier training and testing.

Through this comprehensive processing pipeline, we obtained a feature dataset with the dimensions (4, 61, 40, 2, 70), encompassing multi-dimensional spectral feature data from different sleep stages, providing rich input data for the subsequent deep sleep stage classification task.

4. In order to achieve accurate detection of deep sleep stages, we employed a binary classification strategy framework. We defined the data corresponding to the deep sleep stages as the positive class, and merged the data from other sleep stages (including rapid eye movement (REM), stage 1, and stage 2) as the negative class. Specifically, the spectral features of deep sleep (stages 3 and 4) were treated as positive samples, while those of REM, stage 1, and stage 2 were combined as negative samples. The spectral features from two EEG channels were then summed and averaged,

ensuring that the information from multiple channels was fully utilized for each sample. For the data preparation, we merged the 40 epochs of data from 61 subjects and performed a randomized stratified shuffle, ensuring that the data maintained the correct class distribution. This resulted in a dataset with the shape of (9,760, 70), where there are 9,760 samples, each with 70 frequency features. Given the imbalanced distribution of the data, we adopted Stratified K-Fold cross-validation to partition the data. This method preserves the proportion of positive and negative samples in each fold, ensuring that the test set maintains a similar class distribution as the training set. This approach helps mitigate the impact of class imbalance on model evaluation. The advantage of this partitioning method is that it ensures the fairness of model evaluation and enhances the generalization ability of the classifier. For the power spectral density features obtained, we used a Random Forest classifier for training and classification experiments. Random Forest, as a powerful ensemble learning method, is well-known for its strong noise tolerance and high classification accuracy, making it particularly suitable for high-dimensional feature classification tasks.

In the classification process, we incorporated k-fold cross-validation to conduct multiple rounds of experiments, providing a comprehensive evaluation of the classifier's performance. In each round, a different subset of data was used for training and testing, ensuring the stability and generalization capability of the classifier across various data partitions. Finally, the performance of the method was evaluated based on average classification metrics, such as accuracy, precision, recall, and F1 score, and compared with other common classification methods to verify its superiority.

c. Participants and Experimental Procedure

We used a dataset of size (9,760, 70) for training and testing, where 7,320 samples corresponded to the negative class and 2,440 samples corresponded to the positive class. To evaluate the performance of the classifier, we employed stratified 5-fold cross-validation. This method ensures that the proportion of positive and negative samples in each fold is consistent with the overall dataset, thereby mitigating the impact of class imbalance on model evaluation. Through 5-fold cross-validation, the classifier achieved an average classification accuracy of 95.57%, a recall of 97.35%, and an F1-score of 97.58% across different data subsets. Figure 5 shows the confusion matrix (CM). These results demonstrate the classifier's high precision and recall in accurately identifying deep sleep stages.

For performance evaluation, we used Receiver Operating Characteristic (ROC) curves and Precision-Recall (PR) curves for visual analysis. Figure 7 shows the ROC curve, and Figure 8 shows the PR curve. The ROC curve illustrates the classifier's performance across various decision thresholds, with the Area Under the Curve (AUC) value reaching 99.00%. This high AUC indicates that the model has excellent discriminative power. Additionally, the PR curve shows that the classifier's performance was minimally affected by the class imbalance, maintaining high classification accuracy even in the presence of an uneven distribution of positive and negative samples.

Furthermore, we analyzed the contribution of each feature using feature importance scores derived from the Random Forest classifier. As shown in Figure 9, the most important feature was the Delta wave, which aligns with physiological understanding of deep sleep characteristics.

In addition to classification performance, we conducted a computational cost analysis. The results showed that the average processing time per epoch (30s of EEG data) was approximately 0.06 seconds, which is well below the threshold required for real-time processing. Compared with traditional FFT+SVM and deep learning-based EEGNet models, our method maintained comparable or superior accuracy while achieving lower computational overhead, suggesting better suitability for deployment in real-time or resource-constrained clinical settings.

To avoid potential data leakage, we further compared two validation strategies: 5-fold cross-validation and leave-one-subject-out (LOSO) cross-validation. While 5-fold cross-validation provides an overall assessment of performance stability across different data partitions, LOSO strictly ensures subject independence by preventing data from the same subject appearing in both training and testing sets. The results of LOSO evaluation are presented in Fig. **??**, showing only a slight decrease compared with the original 5-fold results, which demonstrates the strong generalization ability of our method on unseen subjects.

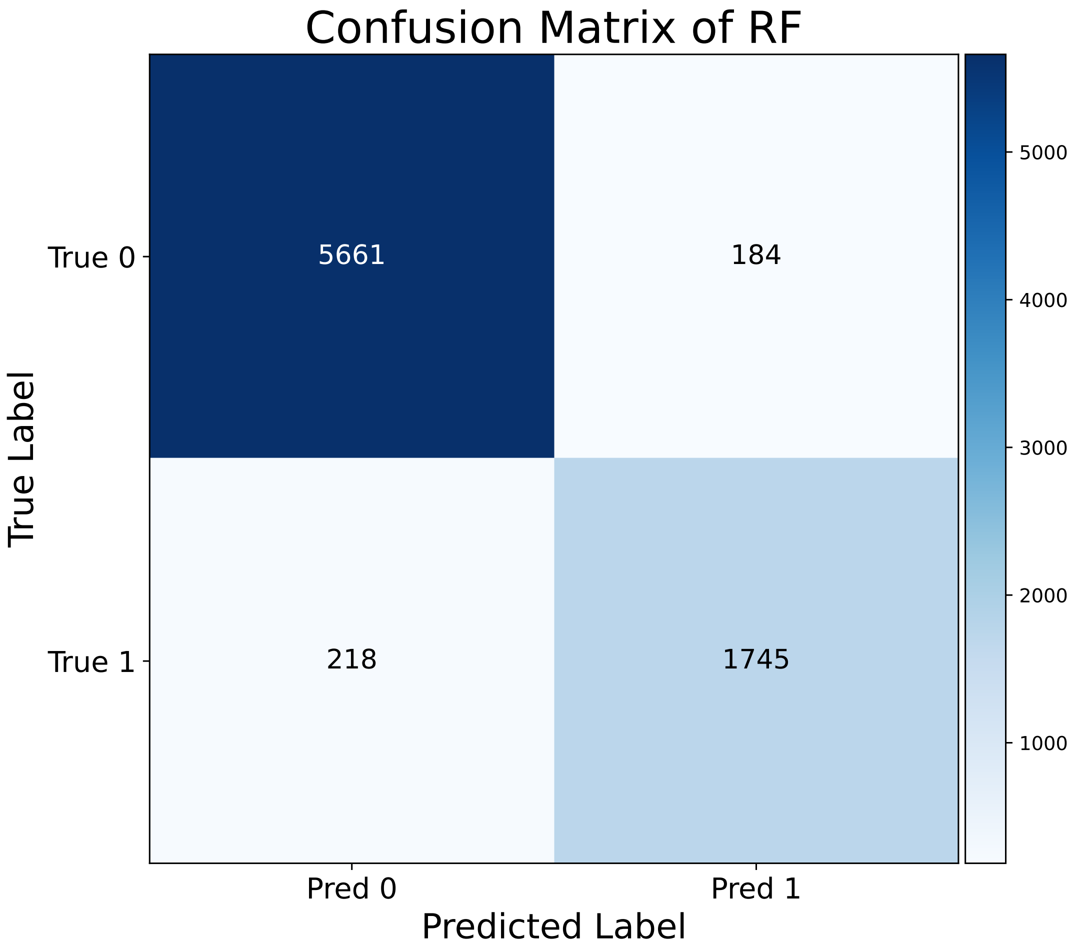


Figure 5: The confusion matrix.

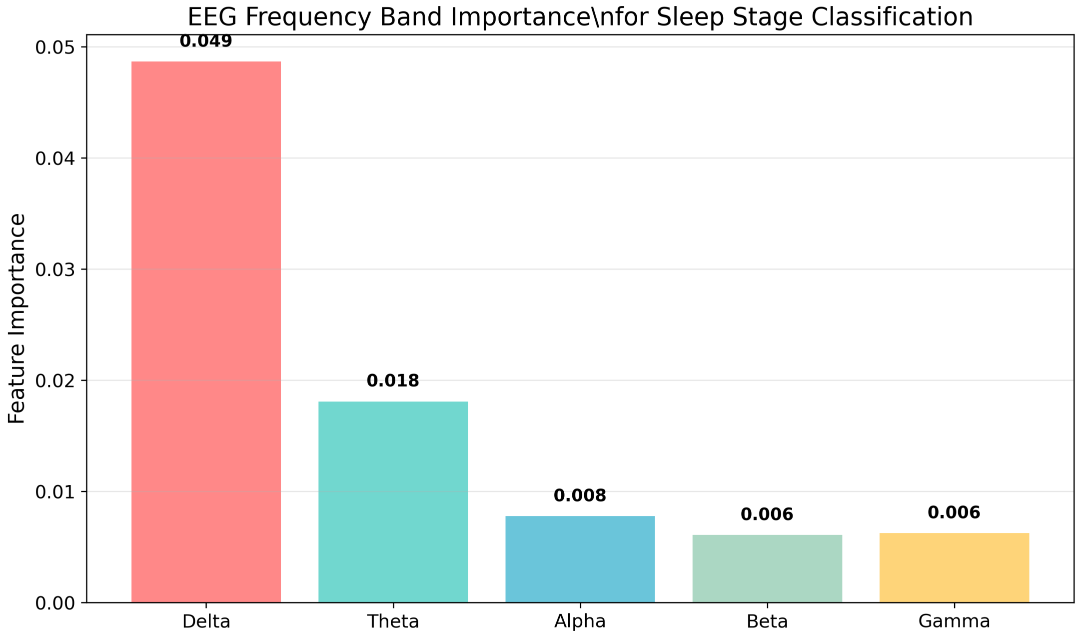


Figure 6: Feature importance analysis.

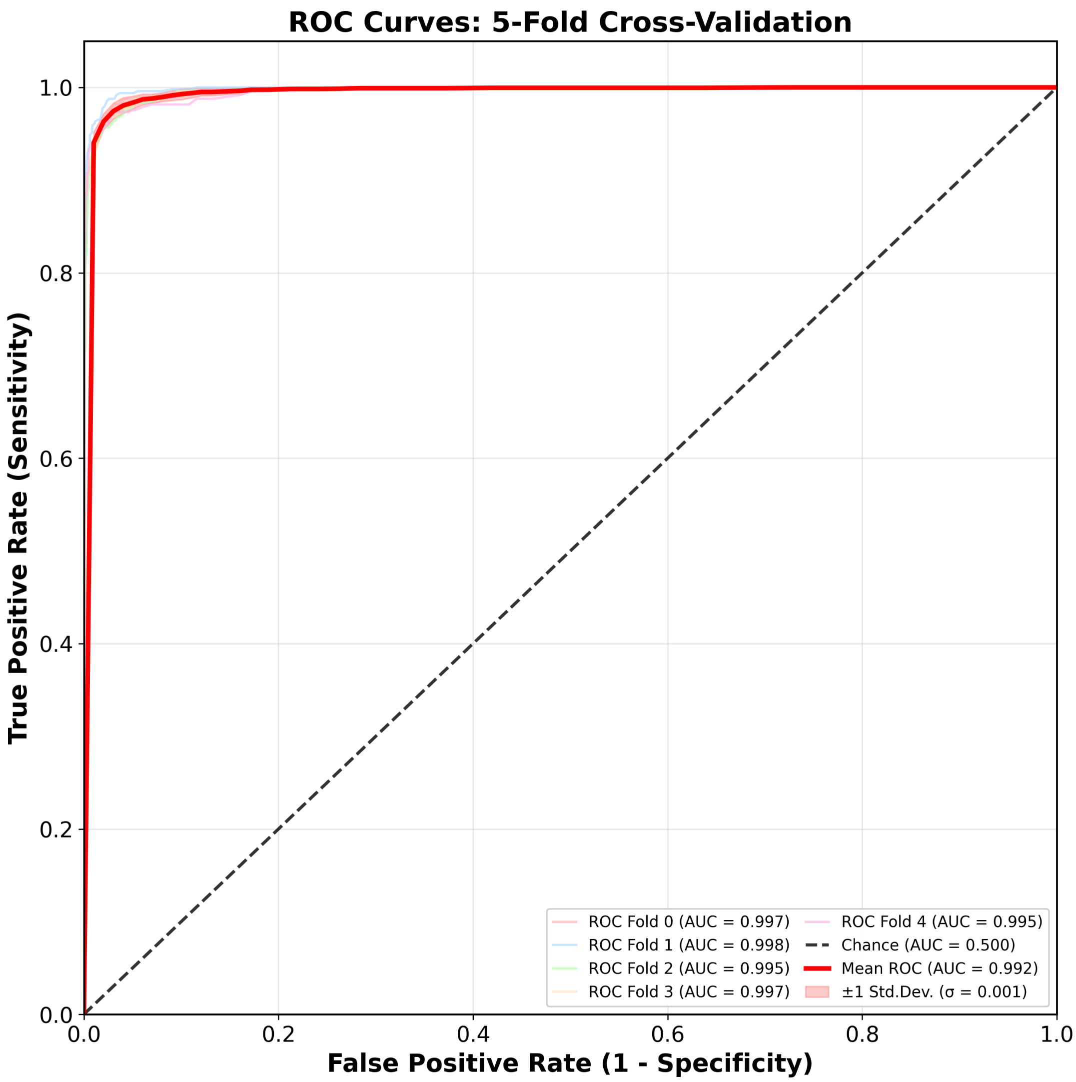


Figure 7: The ROC curve.

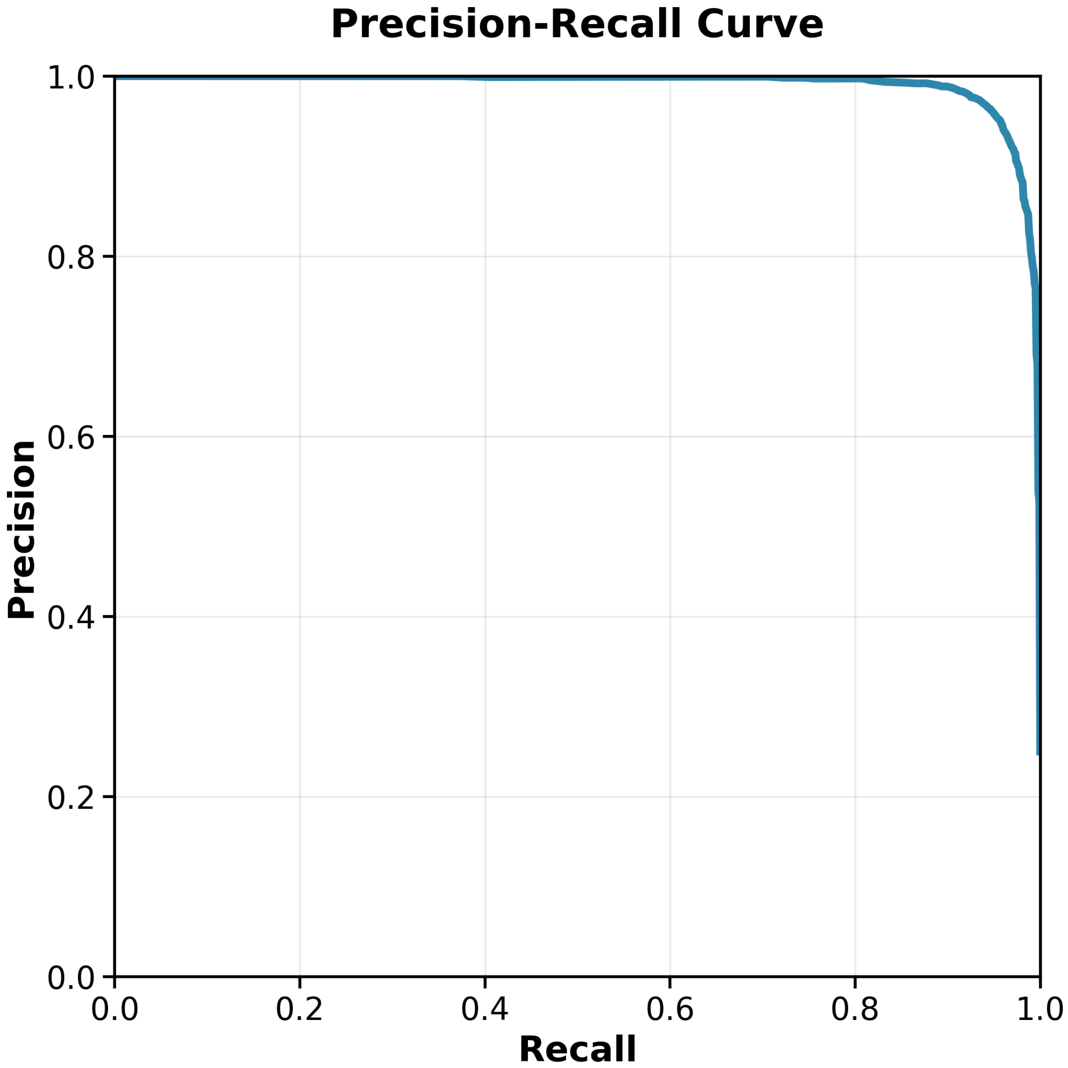


Figure 8: The PR curve.

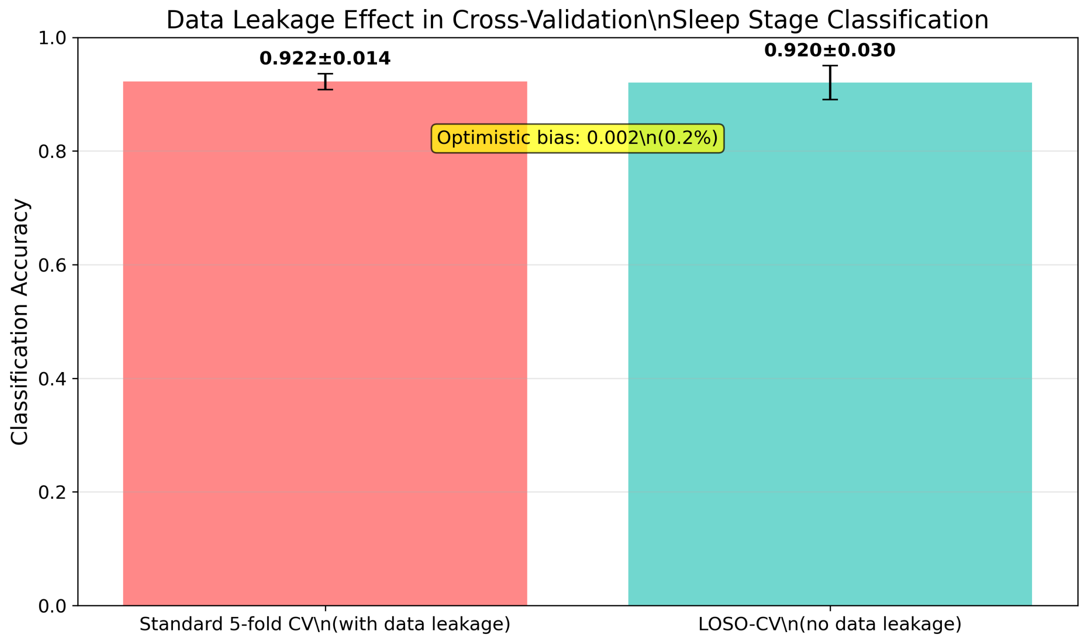


Figure 9: Data leakage comparison.

IV. Discussion

a. Data Processing Considerations

In this study, we followed the standard segmentation protocol of the dataset, using 30-second epochs as the basic unit of analysis. While the 30-second window has been widely adopted in traditional sleep staging research, it may not be the optimal choice from a theoretical standpoint. Longer windows allow for capturing more comprehensive sleep features but may obscure instantaneous sleep state transitions. Conversely, shorter windows are advantageous for capturing rapid changes but may also introduce additional noise. We acknowledge that the use of fixed 30-second windows may obscure temporal transitions, and future research could explore adaptive or dynamic windowing strategies to better reflect the continuous nature of sleep processes. Furthermore, the relatively low sampling frequency of 100 Hz is sufficient for capturing the dominant δ (0.5–4 Hz) and θ (4–8 Hz) bands relevant to deep sleep, but it may limit the characterization of high-frequency activity associated with other sleep stages or neurological conditions. Future studies may consider employing higher sampling rates to obtain richer spectral information.

b. Validation Strategy

The dataset used in this study consisted of EEG recordings from 41 healthy Caucasian subjects and 61 sessions. In principle, leave-one-out cross-validation (LOOCV) could maximize model generalizability. However, we opted for stratified K-fold cross-validation for validation. This decision was made for two reasons: first, with a sufficiently large number of folds, stratified K-fold can approximate the performance of LOOCV; second, stratified K-fold is computationally more efficient, particularly when handling large datasets. This strategy therefore ensures reliable model evaluation while offering computational advantages, making it more suitable for large-scale applications. Nevertheless, the restricted demographic composition of the dataset may limit generalizability to populations with sleep disorders or different ethnic backgrounds. Future work will extend validation to more diverse cohorts to ensure robustness and broad applicability.

c. Computational Complexity and Methodological Advantages

From the perspective of computational efficiency, the proposed method exhibits relatively good performance under our experimental settings. By combining 2T-EMD with MT-PSD, we achieved efficient EEG signal processing. Compared to traditional EMD methods, 2T-EMD ensures mode alignment and suppresses spurious modes without requiring noise injection or extensive ensemble averaging, thereby reducing computational overhead and making it suitable for real-time monitoring. Although MT-PSD is computationally more intensive, it provides stable spectral estimates, and with appropriate optimization it meets real-time processing requirements. For comparison, we also evaluated a lightweight deep learning model (EEGNet). The proposed method achieved comparable or superior accuracy with lower computational cost, highlighting its practical advantage in real-time applications. In future work, incorporating interpretable deep learning strategies, such as CNNs with attention mechanisms, may provide a better trade-off between accuracy and interpretability. This framework therefore strikes a balance between analytical precision and computational efficiency, which is critical for practical deployment in real-time scenarios.

d. Clinical Implications

The results of this study indicate strong potential for clinical application. The high interpretability of the proposed method is particularly advantageous in medical contexts, as it helps clinicians understand and trust the algorithm's diagnostic recommendations. In addition, the method's real-time processing capability makes it suitable for scenarios such as continuous sleep monitoring in clinical or home-based devices, where timely feedback may enable immediate intervention. At the same time, we acknowledge that more comprehensive offline analyses remain valuable for studying complex sleep structures and assessing long-term sleep quality. Nevertheless, practical issues such as environmental noise interference and the stability of long-term EEG recordings need to be addressed before clinical deployment.

e. Limitations and Future Directions

Several limitations of this study should be acknowledged. First, the data selection criterion required participants to have at least 1,500 seconds (50 cycles) of deep sleep. While this ensured adequate representation of the deep sleep stage in the training set, it may also have introduced selection bias. Specifically, this requirement could have excluded individuals with fragmented sleep or poor sleep quality, such as older adults or patients with sleep disorders. Since such populations are not uncommon in clinical practice, the model's performance may decline when applied to these groups. Future research should therefore validate the method in broader and more diverse cohorts to ensure its robustness and generalizability.

Second, while the proposed approach achieved strong performance in a binary classification task (deep sleep vs. other stages), its effectiveness in multi-class problems (e.g., distinguishing REM, N1, and N2 stages) remains to be further validated. We view this as an important direction for future research. Additionally, the current framework primarily emphasizes frequency-domain features; incorporating time-domain information and additional physiological signals could further enhance classification performance. Finally, considering the substantial inter-individual variability in EEG patterns, adaptive or personalized learning mechanisms should be explored to improve model adaptability across diverse populations.

IV. Conclusion

This study presents a novel framework for deep sleep detection using multi-channel EEG signals, integrating 2T-EMD, MT-PSD, and a Random Forest classifier. Validation on 41 subjects (61 recordings) demonstrated high classification accuracy (95.57%) and AUC (99.00%), confirming the method's effectiveness and reliability.

Methodologically, 2T-EMD enables efficient decomposition of multi-channel, non-stationary EEG signals, and in our experimental setup, exhibited favorable computational efficiency. MT-PSD contributes to stable and physiologically meaningful spectral feature extraction. The Random Forest classifier maintains high performance while offering interpretability, which supports clinical insight into the classification outcomes.

The proposed framework shows potential for real-time applications under the tested conditions and demonstrates robust performance across diverse subjects. Future work will focus on optimizing the feature extraction process, extending the framework to other sleep stages, and integrating it with portable monitoring devices, aiming to enhance its clinical applicability and versatility in EEG-based sleep analysis.

a. Ethics approval and consent to participate

The study was approved by the Saitama Institute of Technology Human Research Ethics Committee, conducted under the "Regulations of the Saitama Institute of Technology Human Research Ethics Committee for Studies Involving Human Subjects". Because only publicly available, fully anonymised data from the Sleep-EDF Database Expanded were analysed, the Committee waived further ethical review for this secondary analysis. All participants had provided written informed consent during the original data acquisition.

b. Human ethics

The original data collection complied with all requirements of the Saitama Institute of Technology Human Research Ethics Committee. The present secondary analysis was exempted from additional review.

c. Consent to participate

Written informed consent was obtained from every participant at the time of the original study; no new participants were contacted.

d. Funding

The authors received no financial support for the research, authorship, or publication of this article.

e. Clinical trial registration

Clinical trial registration: not applicable.

f. Data availability

The data analysed are freely available as the Sleep-EDF Database Expanded.

Conflicts of Interest

The author has no conflict of interest about anything in this article.

References

Aboalayon, K. A. I., Faezipour, M., Almuhammadi, W. S., & Moslehpour, S. (2016). Sleep stage classification using eeg signal analysis: A comprehensive survey and new investigation. *Entropy*, *18*(9), 272.

Andrzejak, R. G., Schindler, K., & Rummel, C. (2012). Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients [DOI: https://doi.org/10.1103/PhysRevE.86.046206]. *Physical Review E*, *86*(4), 046206.

Babadi, B., & Brown, E. N. (2014). A review of multitaper spectral analysis. *IEEE Transactions on Biomedical Engineering*, *61*(5), 1555–1564.

Cao, J., & Chen, Z. (2008). Advanced eeg signal processing in brain death diagnosis [DOI: https://doi.org/10.1007/978-0-387-74367-7_15]. In *Signal processing techniques for knowledge extraction and information fusion* (pp. 275–298). Springer.

Chambon, S., Thorey, V., Arnal, P., Mignot, E., & Olivetti, A. H. (2018). A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(4), 758–769. https://doi.org/10.1109/TNSRE.2018.2813138

Chen, Q., Yuan, L., Miao, Y., Zhao, Q., Tanaka, T., & Cao, J. (2019). Quasi-brain-death eeg diagnosis based on tensor train decomposition [DOI: https://doi.org/10.1007/978-3-030-22808-8_49]. *International Symposium on Neural Networks*, 501–511.

Chokroverty, S. (2010). Overview of sleep & sleep disorders. *Indian Journal of Medical Research*, *131*(2), 126–140.

Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C.-K., Li, X., & Guan, C. (2021). An attention-based deep learning approach for sleep stage classification with single-channel eeg [DOI: https://doi.org/10.1109/TNSRE.2021.3076234]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 809–818.

Fattinger, S., de Beukelaar, T. T., Ruddy, K. L., Volk, C., Heyse, N. C., Herbst, J. A., Hahnloser, R. H., Wenderoth, N., & Huber, R. (2017). Deep sleep maintains learning efficiency of the human brain. *Nature communications*, *8*(1), 15405.

Flandrin, P., Rilling, G., & Goncalves, P. (2004). Empirical mode decomposition as a filter bank [DOI: https://10.1109/LSP.2003.821 *IEEE signal processing letters*, *11*(2), 112–114.

Fleureau, J., Nunes, J.-C., Kachenoura, A., Albera, L., & Senhadji, L. (2010). Turning tangent empirical mode decomposition: A framework for mono-and multivariate signals [DOI: https://10.1109/TSP.2010.2097254]. *IEEE Transactions on signal Processing*, *59*(3), 1309–1316.

Gabriel-Alexandru, V., & Mihaela-Ruxandra, L. (2024). Deep learning for sleep staging using eeg signals [DOI: https://10.1109/ISETC63109.2024.10797312]. *2024 International Symposium on Electronics and Telecommunications (ISETC)*, 1–8.

Heckbert, P. (1995). Fourier transforms and the fast fourier transform (fft) algorithm [Lecture Notes, Carnegie Mellon University]. *Computer Graphics*, *2*, 15–463. https://www.cs.cmu.edu/afs/cs/project/anim/ph/463.95/pub/www/notes.toc.html

Jin, Z., & Jia, K. (2023). Sagsleepnet: A deep learning model for sleep staging based on self-attention graph of polysomnography [DOI: https://doi.org/10.1016/j.bspc.2023.105062]. *Biomedical Signal Processing and Control*, *86*, 105062.

Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A., & Oberye, J. J. (2000). Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the eeg [DOI: https://ieeexplore.ieee.org/document/867928]. *IEEE Transactions on Biomedical Engineering*, *47*(9), 1185–1194.

Lozzi, D., Di Matteo, A., Mattei, E., Cipriani, A., Caianiello, P., Mignosi, F., & Placidi, G. (2024). Asis: A smart alarm clock based on deep learning for the safety of night workers [DOI: https://10.1109/MetroXRAINE62247.2024.10796738]. *2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, 1129–1134.

Parmar, A., Katariya, R., & Patel, V. (2018). A review on random forest: An ensemble classifier. *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI 2018)*, 758–763. https://doi.org/10.1007/978-3-030-03146-6_90

Phan, H., Koch, P., Vos, M. D., & Mertins, A. (2021). Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification. *IEEE Transactions on Biomedical Engineering*, *69*(8), 2456–2467. https://doi.org/10.1109/TBME.2021.3094350

Subasi, A., & Gursoy, M. I. (2010). Eeg signal classification using pca, ica, lda and support vector machines. *Expert Systems with Applications*, *37*(12), 8659–8666. https://doi.org/10.1016/j.eswa.2010.06.065

Wang, Y.-H., Yeh, C.-H., Young, H.-W. V., Hu, K., & Lo, M.-T. (2014). On the computational complexity of the empirical mode decomposition algorithm. *Physica A: Statistical Mechanics and its Applications*, *400*, 159–167. https://doi.org/10.1016/j.physa.2014.01.020

Wu, J., Yang, Y., & Wang, X. (2024). Narratingsleep: A language model-based framework for sleep staging [DOI: https://10.1109/BIBM62325.2024.10822537]. *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 3860–3865.

Youngworth, R. N., Gallagher, B. B., & Stamper, B. L. (2005). An overview of power spectral density (psd) calculations. *Optical Manufacturing and Testing VI*, *5869*, 206–216. https://doi.org/10.1117/12.613215

Zhao, X., Zhao, Q., Tanaka, T., Solé-Casals, J., Zhou, G., Mitsuhashi, T., Sugano, H., Yoshida, N., & Cao, J. (2022). Classification of the epileptic seizure onset zone based on partial annotation [DOI: https://doi.org/10.1007/s11571-022-09857-4]. *Cognitive Neurodynamics*, 1–11.