IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 11, No. 12, December, 2024, pp. 3243-3250
Journal homepage: http://iieta.org/journals/mmep

Single Channel Sleep Stage Classification Using Random Forest and Feature Extraction with SMOTE Balancing on Fpz-Cz EEG Data

Kurniawan Eka Permana[1*], Keiji Iramina[2]

[1] Department of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia
[2] Graduate School of Systems Life Sciences, Kyushu University, Fukuoka 819-0395, Japan

Corresponding Author Email: kurniawan@trunojoyo.ac.id

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/mmep.111204

Received: 8 October 2024
Revised: 25 November 2024
Accepted: 2 December 2024
Available online: 31 December 2024

Keywords:
sleep stages classification, Random Forest Classifier, electroencephalography (EEG), single channel EEG, Synthetic Minority Over-sampling Technique (SMOTE)

ABSTRACT

This research presents a novel approach to sleep stage classification using single-channel EEG data and a Random Forest Classifier, integrating advanced feature extraction and SMOTE to address class imbalance. EEG data were preprocessed to extract power band features and time-domain characteristics, such as mean, variance, skewness, kurtosis, and entropy measures (Shannon entropy, permutation entropy, and sample entropy). The study leveraged data from the EEG Fpz-Cz channel to ensure high-quality signal processing, creating epochs and applying a Random Forest model to classify sleep stages into Wake, N1, N2, N3, and REM. SMOTE was used to resample the dataset, ensuring balanced training for the model. The results demonstrated strong performance, with a classification accuracy of 93.5% and a Cohen's Kappa score of 0.92, indicating near-perfect agreement between predicted and actual sleep stages. This study introduces a robust method that simplifies sleep stage analysis by focusing on a single EEG channel, demonstrating its potential for efficient clinical and personal sleep monitoring.

1. INTRODUCTION

Sleep stage classification using EEG data has been a significant area of research due to its implications for understanding sleep patterns and diagnosing sleep disorders. This introduction reviews key studies and methodologies relevant to the classification of sleep stages, with a particular focus on EEG data, feature extraction techniques, and machine learning approaches.

Electroencephalography (EEG) is a widely used method for sleep stage classification due to its ability to capture brain activity patterns associated with different sleep stages. Traditional methods for sleep stage scoring relied on manual annotation of EEG recordings by sleep experts, but recent advancements have shifted towards automated classification using machine learning techniques. Previous researches [1, 2] have shown the effectiveness of various machine learning algorithms, including Random Forests and Support Vector Machines, for automating sleep stage classification. The study highlighted that machine learning models, when trained on comprehensive EEG features, can achieve accuracy comparable to expert annotations [3, 4]. An efficient and scalable solution involves automatic classification of sleep stages using machine learning algorithms with a single EEG channel [5-9]. This approach not only simplifies data acquisition but also reduces computational complexity, making it more accessible for both clinical and home-based applications.

Effective feature extraction is crucial for improving the performance of classification models. EEG data are typically analyzed in both time and frequency domains to extract relevant features. Power Spectrum Density (PSD) is a common frequency-domain feature used to capture oscillatory activities in EEG signals. Rechichi et al. [10] utilized PSD features to classify sleep stages and found that frequency-band analysis significantly enhances classification performance. Additionally, time-domain features such as mean, variance, skewness, and kurtosis have been shown to provide valuable information for distinguishing between sleep stages [11].

Entropy measures are another important feature extraction method. Entropy quantifies the complexity and irregularity of EEG signals, offering insights into the brain's functional state. Shannon entropy, permutation entropy, and sample entropy are frequently used in sleep stage classification. Tripathy et al. [12] demonstrated that incorporating entropy measures improves classification accuracy by capturing the dynamic changes in EEG signals associated with different sleep stages.

Class imbalance is a common issue in sleep stage classification, where certain sleep stages are underrepresented in the dataset. Synthetic Minority Over-sampling Technique (SMOTE) is an effective approach to address this imbalance by generating synthetic samples for minority classes. Salamatian and Khadem [13] applied SMOTE to EEG-based sleep stage classification and reported significant improvements in model performance, particularly for minority sleep stages. SMOTE helps ensure that the classifier is well-trained on all sleep stages, leading to more balanced and accurate predictions.

Machine learning algorithms have proven instrumental across various life sciences applications, including sleep research. Techniques such as Random Forest [14], Convolutional Neural Networks (CNNs) [15], Neural Networks (NNs) [16], and fuzzy logic systems [17] are commonly employed to improve classification accuracy. Random Forests are a popular choice for sleep stage classification due to their robustness and ability to handle high-dimensional data. The Random Forest algorithm, as demonstrated by Sundararajan et al. [14], can effectively classify sleep stages by combining multiple decision trees to make predictions. Random Forests offer advantages such as reduced overfitting and high accuracy, making them suitable for complex classification tasks.

In addition to Random Forests, other machine learning algorithms such as Neural Networks and Gradient Boosting have also been explored in sleep stage classification. For instance, a study by Satapathy et al. [18] compared various machine learning models and found that Gradient Boosting provided competitive performance with Random Forests. However, the choice of model often depends on the specific characteristics of the dataset and the features used.

While previous studies have effectively utilized machine learning models like Random Forests and SVMs for multi-channel EEG data [1, 2], they often involve complex setups, limiting accessibility for home-based or portable applications. Single-channel approaches, as explored by Zhou et al. [5] and Nguyen et al. [6], simplify data acquisition but require robust methodologies to maintain accuracy. Feature extraction methods such as PSD and time-domain metrics are well-established for identifying oscillatory patterns [10]; however, these approaches may underperform in imbalanced datasets. Our study leverages SMOTE to address this limitation, as demonstrated by Salamatian and Khadem [13], where class balancing improved model performance significantly. By integrating entropy measures into feature extraction, our approach captures dynamic EEG signal complexities, a method less frequently addressed in prior works [12].

Despite progress in sleep stage classification, there remains a gap in achieving high accuracy with single-channel EEG data while addressing class imbalance effectively. This study addresses this gap by combining comprehensive feature extraction with SMOTE and Random Forest to answer the research question: Can single-channel EEG data with enhanced feature extraction and class balancing achieve comparable accuracy to multi-channel setups in sleep stage classification?

The primary objective of this study is to demonstrate that a simplified, single-channel EEG framework can achieve high accuracy and reliability in sleep stage classification through advanced feature extraction and class balancing techniques, paving the way for accessible sleep analysis solutions.

In this research, we focus on sleep stage classification using a single EEG channel, specifically the "EEG Fpz-Cz" channel, a common placement in sleep studies [5, 8, 9]. By leveraging the Random Forest Classifier, we aim to accurately classify the five key sleep stages: wake (W), non-rapid eye movement (NREM) stages N1, N2, N3, and rapid eye movement (REM) sleep. Random Forest (RF) was chosen as the model for sleep stage classification due to its inherent advantages in handling high-dimensional data, robustness against overfitting, and ability to capture complex patterns in EEG features. RF is particularly effective for datasets with mixed features, such as time-domain statistics, frequency-domain power bands, and entropy measures, as it can naturally handle heterogeneous input types. Its ensemble nature, where multiple decision trees are combined, increases its predictive stability and accuracy compared to single decision-tree classifiers. Compared to other classifiers like Support Vector Machines (SVMs) or neural networks, RF requires minimal tuning and is less sensitive to hyperparameters, making it a practical choice for an initial baseline model. Unlike neural networks, RF is less computationally intensive and does not require extensive preprocessing or scaling of features.

To further enhance the classification accuracy, particularly in handling imbalanced data common in sleep studies, we apply Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates synthetic samples of under-represented sleep stages, ensuring that the Random Forest model is trained on a more balanced dataset. Additionally, a combination of frequency and time-domain features, including EEG power bands and entropy measures, is extracted to provide the classifier with a comprehensive feature set.

This research aims to demonstrate that a single-channel EEG, combined with advanced feature extraction techniques and a balanced dataset, can achieve high accuracy in sleep stage classification. The study also highlights the impact of SMOTE in mitigating the bias introduced by imbalanced sleep stage distributions. Through this work, we seek to contribute to the development of more accessible and efficient tools for sleep monitoring and analysis, with potential applications in both clinical and personal health contexts.

2. METHODOLOGY

The methodology of this research can be seen in Figure 1, involves several key steps to classify sleep stages using EEG data with a Random Forest Classifier, enhanced by feature extraction and SMOTE for balancing.

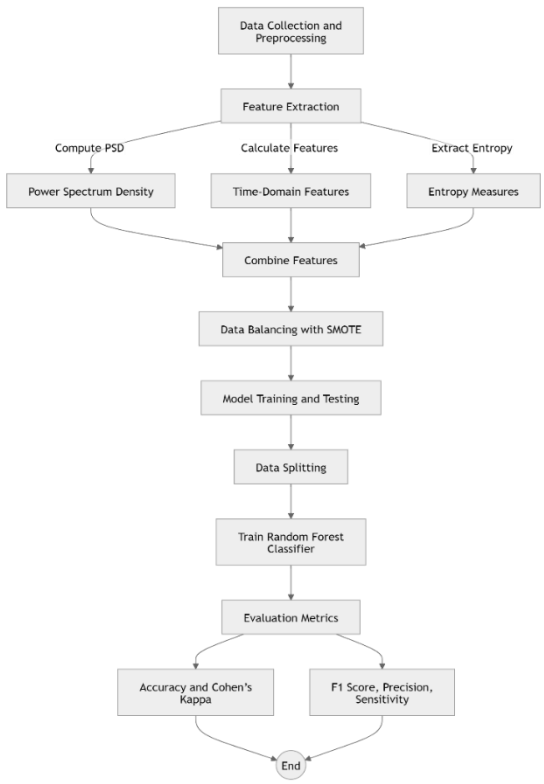


Figure 1. Research methodology of RF-SMOTE

2.1 Data collection and preprocessing

Data Collection: The dataset utilized in this work is the second adaptation of the PhysioNet Sleep-EDF dataset [19], which was developed in 2018 and includes 197 polysomnograms (PSG) to assess the proposed model's performance for sleep stage assignment. The Sleep-EDF dataset includes two types of data: (1) those that assess the effects of age on sleep in healthy people, and (2) those that examine the effects of temazepam (sleeping pills) on sleep. Eight specific datasets, four from each type, are chosen. Professionals physically labeled the hypnograms using Rechtschaffen and Kales standards, assigning each to a separate class at each level. The American Academy of Sleep Medicine (AASM) defines these classes as W, REM, N1, N2, and N3 [20]. In the evaluations, EEG data from single channels of both versions were integrated for analysis. EEG data were collected from the Fpz-Cz channel of EEG recordings. Files containing both EEG data (PSG) and annotations (Hypnogram) were used. These files were loaded from a specified directory, and preprocessing was applied.

Preprocessing: The raw EEG data were segmented into epochs based on annotated sleep stages. Epochs were created with a duration of 30 seconds, and the data were prepared for further analysis. The signal undergoes bandpass filtering between 0.5 Hz and 30 Hz to retain relevant sleep-related frequencies while removing noise and artifacts.

2.2 Feature extraction

Feature extraction involves calculating both frequency-domain and time-domain features.

Power Spectrum Density (PSD): The frequency-domain features include relative power for the delta (0.5–4.5 Hz), theta (4.5–8.5 Hz), alpha (8.5–11.5 Hz), sigma (11.5–15.5 Hz), and beta (15.5–30 Hz) bands, computed from the PSDs. These bands capture various oscillatory activities relevant to sleep stages. The Power Spectrum Density (PSD) is calculated as:

$$P_x(f) = \frac{|FFT(x(t))|^2}{N} \tag{1}$$

where, N is the window length and FFT is the Fast Fourier Transform.

The power in each band was normalized to ensure comparability.

Time-Domain Features: Features such as mean, variance, skewness, and kurtosis of the EEG signal were calculated for each epoch.

Sleep stage classification using EEG data has been a significant area of research due to its implications for understanding sleep patterns and diagnosing sleep disorders. This literature review explores key studies and methodologies relevant to the classification of sleep stages, particularly focusing on the use of EEG data, feature extraction techniques, and machine learning approaches.

$$Mean = \frac{\sum_{i=1}^{N} X_i}{N} \tag{2}$$

$$\sigma^2(Variance) = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N} \tag{3}$$

$$Skew = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^3 \tag{4}$$

$$Kurt = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^4 \tag{5}$$

where, $\bar{X}$ is the mean and σ is the standard deviation.

Entropy Measures: Entropy features were extracted to capture the complexity of the EEG signal. This included Shannon entropy, permutation entropy, and sample entropy. These measures provide insights into the randomness and structure of the EEG data. Shannon entropy is calculated as:

$$H(x) = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i) \tag{6}$$

where, $p(x_i)$ is the probability of the i-th state.

Permutation entropy is calculated as:

$$\mathrm{PE}_\tau^D(X) = \sum_{i=1}^{D!} \frac{-p_\tau^D(\pi_i)\ln(p_\tau^D(\pi_i))}{\ln D!} \tag{7}$$

where, $p_\tau^D(\pi_i)$ is the probability distribution of the i-th permutation and D is the embedding dimension.

Sample entropy (SampEn) is calculated as:

$$SampEn = -\log\left(\frac{(\sum A_i)}{(\sum B_i)}\right) = -log\frac{A}{B} \tag{8}$$

where, A and B are the number of matches in two different datasets of length m.

These features were chosen as they are highly informative for distinguishing sleep stages, with spectral features linked to physiological rhythms (e.g., alpha for wakefulness) and entropy capturing dynamic signal variations. The features from different domains are concatenated into a single feature vector for each epoch of data.

2.3 Data balancing with SMOTE

The dataset was balanced using Synthetic Minority Over-sampling Technique (SMOTE) to address the issue of class imbalance [21]. SMOTE was applied after feature extraction but before splitting the dataset into training and testing sets. This ensured that synthetic samples were generated only from the training set to prevent data leakage. SMOTE generates synthetic samples by interpolating between existing minority class samples, preserving class characteristics and avoiding simple duplication of samples. Alternative techniques such as random oversampling risk overfitting, while undersampling can result in loss of valuable information from the majority class.

Input:
Minority class samples X_min
Number of nearest neighbors k
Desired number of synthetic samples N
For each sample x in X_min:
1. Find k-nearest neighbors of x in X_min
2. Randomly select a neighbor x_neighbor from the k-nearest neighbors
3. Generate a new synthetic sample:
diff = x_neighbor - x
*new_sample = x + r * diff, where r is a random number between 0 and 1*
4. Add new_sample to the dataset
Repeat until N synthetic samples are generated.
Output: Augmented dataset with new synthetic minority class samples

---

2.4 Model training and testing

Data Splitting: The train-test split was randomized, with 20% of the data allocated for testing and 80% for training. Stratification was applied during the split to maintain the distribution of sleep stages across training and testing datasets, ensuring that minority classes were adequately represented in both sets. While no explicit cross-validation was performed in this experiment, future work could incorporate techniques like k-fold or stratified k-fold cross-validation to better assess model generalizability and reduce potential biases in the evaluation process.

Random Forest Classifier: An ensemble Random Forest model, trained on the extracted features [22, 23], was used for classification. This model was chosen for its robustness and ability to handle complex data.

Input:
Dataset D = {(x1, y1), (x2, y2), ..., (xn, yn)}
Number of trees T
Number of features to consider per split m

For t = 1 to T do:
1. Draw a bootstrap sample from D
2. Grow a decision tree on this bootstrap sample:
For each node in the tree:
a. Randomly select m features from the total features
b. Split the node on the feature that results in the best split
Continue until the maximum depth is reached or no further split is possible

End For

For each new data point x:
1. Send x through each of the T trees to obtain predictions
2. Use majority voting to assign the final predicted class

Output: Final prediction for all data points

The Random Forest model was designed to optimize classification accuracy while minimizing overfitting. Key parameters were configured as follows: Number of Trees (n estimators): 100, maximum Depth: None, split Criterion: Gini impurity and minimum Samples per Split: 2.

2.5 Evaluation metrics

Model performance was evaluated using accuracy and Cohen's Kappa score [24]. Accuracy measures the proportion of correct predictions, while Cohen's Kappa assesses the agreement between predicted and true labels, considering chance.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} = \frac{\sum_{i=1}^{N} I(y_i, \acute{y}_i)}{N} \quad (9)$$

$$\kappa(Cohen's\ Kappa) = \frac{P_o - P_e}{1 - P_e} \quad (10)$$

where, P_o is the observed agreement, which is the proportion of instances where both raters agree; P_e is the expected agreement by chance, calculated based on the marginal probabilities of each rater.

Additional metrics such as F1 Score, precision, and sensitivity were calculated for each sleep stage to assess the model's performance in identifying each stage accurately [25].

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \quad (12)$$

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \quad (13)$$

3. RESULT AND DISCUSSION

In this research, we focus on sleep stage classification using a single EEG channel, specifically utilizing the Random Forest Classifier and SMOTE (RF-Smote). The single-channel EEG data provides an efficient and accessible alternative to multi-channel systems, reducing complexity while maintaining accuracy. Random Forest, a robust ensemble learning algorithm, is employed to classify the five major sleep stages: wake (W), NREM stages N1, N2, N3, and REM sleep. SMOTE is applied to address the common issue of class imbalance in sleep datasets by generating synthetic samples for under-represented stages, thereby improving the model's performance and accuracy across all sleep stages. This combination aims to enhance classification accuracy while maintaining simplicity in data acquisition.

Figure 2 highlights the dataset's class imbalance, with the Wake stage (W) significantly overrepresented compared to other stages. N1 has the fewest samples, while N2, N3, and REM (R) are moderately represented but still underrepresented relative to W. This imbalance can bias the model toward overpredicting the majority class (W) and underperforming on minority classes like N1 and N3, leading to misleading overall accuracy that does not reflect true performance across all sleep stages [26].

After applying SMOTE, each sleep stage will have 8037 data points, balancing the dataset to match the highest class count of stage W. Applying SMOTE will help balance the dataset by generating synthetic data for the minority classes, improving the model's ability to learn from all classes equally.

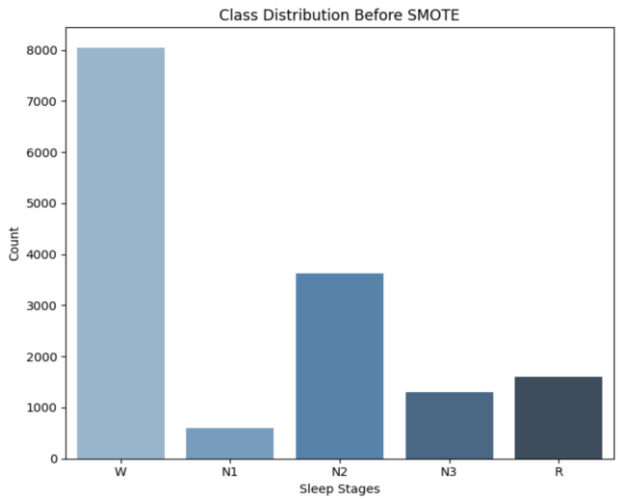


Figure 2. Class distribution before SMOTE

Table 1. Performance of RF-SMOTE

Metric	R	N3	N2	N1	W
F1-score	0.926721	0.951479	0.898718	0.920895	0.974343
Precision	0.9338	0.939252	0.91395	0.905766	0.980709
Sensitivity	0.919749	0.964029	0.883985	0.936538	0.968059

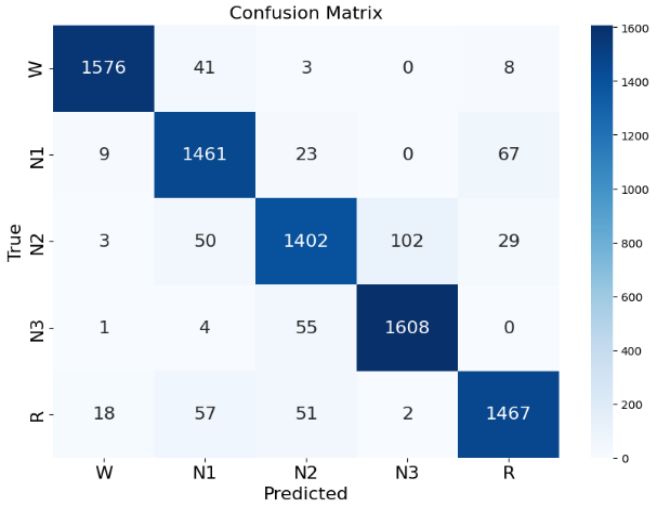


Figure 3. Confusion matrix of RF-SMOTE model

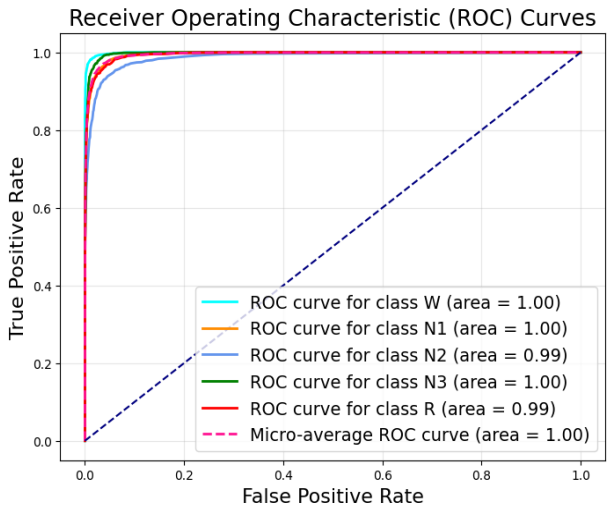


Figure 4. RUC Curve of RF-SMOTE model

The results from Table 1 demonstrate the RF-Smote model's strong performance, particularly for Wake and N3 stages. The model achieves the highest Precision for Wake at 0.981, indicating excellent accuracy in predicting this stage. N3 follows with a Precision of 0.939. The F1-score for Wake is 0.974, the highest among all stages, while N3 has an F1-score of 0.952, both reflecting well-balanced performance. REM has an F1-score of 0.927, and N1 scores 0.921, showing good but slightly lower performance. N2 has the lowest F1-score at 0.899, indicating some difficulty in predicting this stage. Sensitivity is highest for N3 at 0.964, followed by Wake at 0.968, showing strong detection capabilities for these stages. Sensitivity for REM and N1 are 0.920 and 0.937, respectively, while N2 has the lowest at 0.884, indicating more frequent misclassifications for Stage 2.

Figure 3 shows the confusion matrix for the RF-Smote model, highlighting the key performance of the model in predicting sleep stages. The model excels in correctly identifying Wake, with 1576 accurate predictions, though it misclassified a few as Stage 1 (41), Stage 2 (3), and REM (8). For Stage 1 (N1), the model accurately predicted 1461 instances but misclassified some as Wake (9), Stage 2 (23), and REM (67). Stage 2 (N2) was correctly predicted 1402 times, but there were misclassifications, notably 102 as N3 and 29 as REM. N3 was accurately predicted 1608 times, with minor misclassifications as Stage 1 (4) and Stage 2 (55). REM sleep (R) was correctly predicted 1467 times, but it was misclassified as Wake (18), Stage 1 (57), and Stage 2 (51). The model shows strong performance in predicting Wake and N3, with more confusion occurring between Stage 2 and other stages like Stage 1 and REM.

Figure 4 display RUC curve that illustrates the performance of a classifier in distinguishing between different sleep stages. The curves show how well the classifier performs for each individual sleep stage, with each line corresponding to a different sleep stage, such as W, N1, N2, N3, and R. These curves plot the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold values.

Each sleep stage has an AUC (Area Under the Curve) close to or equal to 1.00, indicating that the classifier performs excellently across all stages. AUC values close to 1.00 signify that the model does a great job of correctly classifying sleep stages with very few errors.

The dashed pink line represents the micro-average ROC curve, which aggregates the contributions of all sleep stages into a single curve. The AUC of the micro-average curve is also 1.00, further suggesting that the overall classification performance is highly accurate.

The blue dashed diagonal line represents the baseline of random guessing, where a random classifier would perform. Since the actual ROC curves for each sleep stage are far above this baseline, it shows that the classifier is doing much better than random guessing.

Figure 5 compares the true and predicted hypnograms, showing the model's ability to accurately predict sleep stages based on EEG features. A close match between the true and predicted hypnograms indicates strong model performance, as it demonstrates correct transitions between stages like Wake, REM, and different NREM stages. The true hypnogram represents the actual sleep stages labeled in the dataset, providing the ground truth for evaluation. The predicted hypnogram displays the stages as the model predicts them, ideally mirroring the true hypnogram. Small differences between the hypnograms suggest the model is mostly accurate in predicting sleep stages, though it may struggle to distinguish between stages with similar EEG patterns, such as N1 and N2.

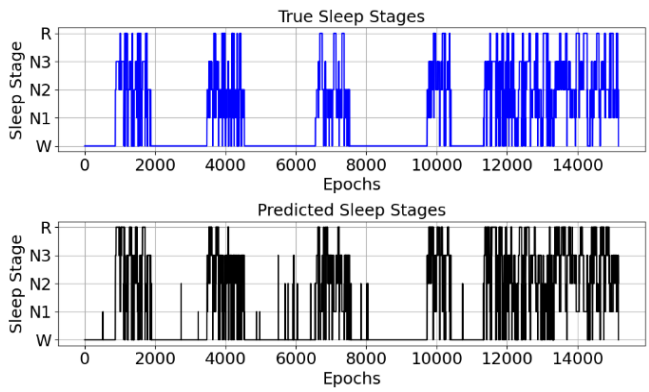


Figure 5. Comparison between hypnogram of true and predicted results

Table 2. Performance comparison with difference classification method

Model	Accuracy	Cohen Kappa
NN	0.85234	0.758344
GB	0.860271	0.825307
SVM	0.877719	0.801346
RF	0.899802	0.838389
RF Smote	0.934926	0.918645

Table 2 highlights the performance of five classification models—Neural Network (NN), Gradient Boosting (GB), Support Vector Machine (SVM), Random Forest (RF), and Random Forest with SMOTE (RF Smote)—in terms of Accuracy and Cohen's Kappa, key metrics for evaluating classification effectiveness.

The RF Smote model stands out with the highest Accuracy (93.5%) and Cohen's Kappa (0.918), demonstrating excellent performance and the ability to handle class imbalance effectively. This approach ensures consistent and reliable predictions across all sleep stages. The standard RF model follows with an Accuracy of 90% and a Cohen's Kappa of 0.838, indicating robust predictive capabilities and strong agreement with true labels. SVM achieves an Accuracy of 87.8%, slightly outperforming GB (86%) and NN (85.2%). However, its Cohen's Kappa of 0.801 is lower than GB's 0.825, suggesting that while SVM has high accuracy, GB offers better agreement when accounting for chance. NN lags behind, with the lowest Accuracy (85.2%) and Cohen's Kappa (0.758), reflecting relatively weaker performance in this context.

Overall, the RF Smote model delivers the most reliable and accurate classification, setting a benchmark for handling imbalanced datasets effectively.

The performance of the model for sleep stage classification demonstrates promising results but also highlights areas for improvement. The model's high F1-scores for Wake (W) and N3, with values of 0.974 and 0.952, respectively, suggest that these stages, which exhibit distinctive EEG patterns, are well-suited to classification by machine learning algorithms. These findings are consistent with previous work by Wen [1], who also observed high accuracy in detecting Wake and N3 stages using EEG data. The clear EEG signatures during these stages, such as mixed-frequency activity in wakefulness and delta waves in, make them easier for the model to identify. Figure 1 shows the model's performance for these stages, where the classifier demonstrates a clear separation between the two based on EEG features.

On the other hand, the model showed some difficulty with Stage 2 (N2) classification, where the F1-score was 0.899. Misclassifications in N2 were primarily between N2 and Stage 1 (N1), reflecting the similarity in their low-frequency components. This issue is well documented in the literature, where the transition between N1 and N2 often results in confusion, especially when the sleep cycle is disturbed or there is artifact noise in the signal [2]. Figure 2 illustrates this problem, showing the confusion between N1 and N2, where a large portion of N2 samples were incorrectly labeled as N1. This misclassification can be particularly problematic in clinical settings where accurate staging of light sleep is essential for diagnosing sleep disorders. A similar challenge was noted by Zhuang et al. [3], who found that automated sleep stage classification systems often struggle with the N1-N2 boundary.

The misclassification of REM (R) sleep was also notable, with many R stages being incorrectly classified as N2 or N1. This issue can be attributed to the similarities in the frequency bands during REM and light sleep stages, where both can exhibit mixed-frequency activity, as discussed by Zhou et al. [5]. Figure 3 demonstrates the confusion matrix for REM, showing that it is frequently confused with N2. The underlying cause is likely due to the shared characteristics of these stages, which may need more specific features to distinguish them effectively.

The model's performance could be improved by incorporating more advanced features that better capture the temporal and spectral nuances of these stages. As suggested by Nguyen et al. [6], multi-scale entropy or wavelet transforms could offer better resolution in detecting subtle transitions between sleep stages, particularly in light sleep and REM stages. These methods are known to enhance classification accuracy by focusing on different time scales, which could help differentiate between stages like N1, N2, and REM, which share similar EEG patterns.

Despite the promising results, several limitations should be acknowledged. The use of SMOTE for class balancing helped address some of the class imbalances, but misclassifications in the minority stages (N1, N2, and REM) still occurred. This reflects the inherent difficulty of classifying these stages accurately in the presence of noise or subtle signal transitions. Further exploration of advanced synthetic data generation techniques, such as Generative Adversarial Networks (GANs), could help create more representative samples of underrepresented stages [21].

Although single-channel EEG data is useful for practical applications, it provides limited spatial information compared to multi-channel EEG. As seen in previous studies [7], multi-channel data allows for the extraction of more detailed features related to brain activity during sleep. Future research should consider multi-channel systems, as they could significantly enhance classification performance, particularly for more difficult stages like N1 and N2.

The model's feature set could be expanded to include more advanced spectral analysis methods, such as wavelet transforms or time-frequency representations, which have been shown to improve classification accuracy for sleep staging tasks [12]. Additionally, integrating multi-modal signals, such as ECG or respiration data, could help improve the robustness of the classifier, especially for distinguishing between stages like N1, N2, and REM [6].

While Random Forest was effective, exploring more complex models, such as deep learning approaches (e.g., CNNs or LSTMs), could further improve classification accuracy. These models have shown success in other sleep stage classification tasks due to their ability to capture temporal dependencies in EEG signals [13]. Investigating these models could help improve performance, especially for more challenging sleep stages like REM.

4. CONCLUSIONS

The study demonstrates high accuracy in sleep stage classification, with a model achieving 93.5% accuracy and a Cohen's Kappa score of 0.92, indicating solid prediction performance. The use of SMOTE to address class imbalance effectively improved the model's ability to learn from all classes, enhancing its performance on minority classes like N3 and REM. The comprehensive EEG feature extraction, which

includes power spectrum, time-domain statistics, and entropy measures, provides a deep understanding of EEG signal characteristics, contributing to accurate classification. Additionally, the use of a single EEG channel simplifies data collection and analysis, reducing costs and complexity compared to multiple channels.

However, the study also has some limitations. The model shows lower performance in classifying lighter sleep stages, especially N2, with reduced F1-score and sensitivity. This indicates that the model struggles to distinguish this stage from others. The reliance on a single EEG channel, while reducing complexity, may not capture all relevant features of the EEG signal compared to using multiple channels. Although SMOTE helps manage class imbalance, the original data still exhibits uneven class distribution, which can impact the model's performance if not properly addressed. Finally, while the model performs well on the current dataset, it may not generalize easily to other datasets or different populations without further validation.

Overall, while the research presents a strong foundation for sleep stage classification, there are areas for improvement and considerations for broader application.

ACKNOWLEDGMENT

I would like to express my deepest gratitude to my advisors and mentors for their invaluable guidance, encouragement, and constructive feedback throughout this research. I am also thankful to my colleagues and peers for their support and collaboration, which greatly enriched this work.

REFERENCES

[1] Wen, W. (2021). Sleep quality detection based on EEG signals using transfer support vector machine algorithm. Frontiers in Neuroscience, 15: 670745. https://doi.org/10.3389/fnins.2021.670745

[2] Liu, J., Wu, D., Wang, Z., Jin, X., Dong, F., Jiang, L., Cai, C. (2020). Automatic sleep staging algorithm based on random forest and hidden Markov model. Computer Modeling in Engineering & Sciences, 123(1): 401-426. https://doi.org/10.32604/cmes.2020.08731

[3] Zhuang, D., Rao, I., Ibrahim, A.K. (2022). A machine learning approach to automatic classification of eight sleep disorders. arXiv preprint arXiv:2204.06997. https://doi.org/10.48550/arXiv.2204.06997

[4] Houssein, E.H., Hammad, A., Ali, A.A. (2022). Human emotion recognition from EEG-based brain–computer interface using machine learning: A comprehensive review. Neural Computing and Applications, 34(15): 12527-12557. https://doi.org/10.1007/s00521-022-07292-4

[5] Zhou, D., Wang, J., Hu, G., Zhang, J., Li, F., Yan, R., Kettunen, L., Chang, Z., Xu, Q., Cong, F. (2022). SingleChannelNet: A model for automatic sleep stage classification with raw single-channel EEG. Biomedical Signal Processing and Control, 75: 103592. https://doi.org/10.1101/2020.09.21.306597

[6] Nguyen, A.T., Nguyen, T., Le, H.K., Pham, H.H., Do, C. (2022). A novel deep learning-based approach for sleep apnea detection using single-lead ECG signals. In 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 2046-2052. https://doi.org/10.23919/APSIPAASC55919.2022.9979890

[7] Permana, K.E., Okamoto, T., Iramina, K. (2018). Single channel electroencephalogram measurement with multi-scale entropy analysis for evaluating day time sleep. In 6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6), pp. 431-435. https://doi.org/10.1007/978-981-10-4361-1_73

[8] Berthomier, C., Drouot, X., Herman-Stoïca, M., Berthomier, P., Prado, J., Bokar-Thire, D., Benoit, O., Mattout, J., d'Ortho, M.P. (2007). Automatic analysis of single-channel sleep EEG: Validation in healthy individuals. Sleep, 30(11): 1587-1595. https://doi.org/10.1093/sleep/30.11.1587

[9] Zhao, S., Long, F., Wei, X., Ni, X., Wang, H., Wei, B. (2022). Evaluation of a single-channel EEG-based sleep staging algorithm. International Journal of Environmental Research and Public Health, 19(5): 2845. https://doi.org/10.3390/ijerph19052845

[10] Rechichi, I., Zibetti, M., Borzì, L., Olmo, G., Lopiano, L. (2021). Single-channel EEG classification of sleep stages based on REM microstructure. Healthcare Technology Letters, 8(3): 58-65. https://doi.org/10.1049/htl2.12007

[11] Jia, Z., Lin, Y., Wang, J., Ning, X., He, Y., Zhou, R., Zhou, Y., Lehman, L.W.H. (2021). Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29: 1977-1986. https://doi.org/10.1109/TNSRE.2021.3110665

[12] Tripathy, R.K., Ghosh, S.K., Gajbhiye, P., Acharya, U.R. (2020). Development of automated sleep stage classification system using multivariate projection-based fixed boundary empirical wavelet transform and entropy features extracted from multichannel EEG signals. Entropy, 22(10): 1141. https://doi.org/10.3390/e22101141

[13] Salamatian, A., Khadem, A. (2020). Automatic sleep stage classification using 1D convolutional neural network. Frontiers in Biomedical Technologies, 7(3): 142-150. https://doi.org/10.18502/fbt.v7i3.4616

[14] Sundararajan, K., Georgievska, S., Te Lindert, B.H., Gehrman, P.R., Ramautar, J., Mazzotti, D.R., Sabia, S., Weedon, M.N., van Someren, E.J.W., Ridder, L., Wamg, J., van Hees, V.T. (2021). Sleep classification from wrist-worn accelerometer data using random forests. Scientific Reports, 11(1): 24. https://doi.org/10.1038/s41598-020-79217-x

[15] Hamdi, M., Inan, T. (2023). Enhanced emotion recognition through the integration of gated recurrent unit and convolutional neural networks using MindWave mobile EEG device. Mathematical Modelling of Engineering Problems, 10(5): 1643-1656. https://doi.org/10.18280/mmep.100514

[16] Ramdass, P., Ganesan, G. (2023). Leveraging neighbourhood component analysis for optimizing multilayer feed-forward neural networks in heart disease prediction. Mathematical Modelling of Engineering Problems, 10(4): 1317-1323. https://doi.org/10.18280/mmep.100425

[17] Mahdi, H.A., Shujaa, M.I., Zghair, E.M. (2023). Diagnosis of medical images using Fuzzy Convolutional

Neural Networks. Mathematical Modelling of Engineering Problems, 10(4): 1345-1351. https://doi.org/10.18280/mmep.100428

[18] Satapathy, S.K., Bhoi, A.K., Loganathan, D., Khandelwal, B., Barsocchi, P. (2021). Machine learning with ensemble stacking model for automated sleep staging using dual-channel EEG signal. Biomedical Signal Processing and Control, 69: 102898. https://doi.org/10.1016/j.bspc.2021.102898

[19] Sundar, G.N., Narmadha, D., Jone, A.A.A., Sagayam, K.M., Dang, H., Pomplun, M. (2021). Automated sleep stage classification in sleep Apnoea using convolutional neural networks. Informatics in Medicine Unlocked, 26: 100724. https://doi.org/10.1016/j.imu.2021.100724

[20] Berry, R.B., Gamaldo, C.E., Harding, S.M., Brooks, R., Lloyd, R.M., Vaughn, B.V., Marcus, C.L. (2015). AASM scoring manual version 2.2 updates: New chapters for scoring infant sleep staging and home sleep apnea testing. Journal of Clinical Sleep Medicine, 11(11): 1253-1254.

[21] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16: 321-357. https://doi.org/10.1613/jair.953

[22] Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a Random Forest Classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67: 93-104. https://doi.org/10.1016/j.isprsjprs.2011.11.002.

[23] Smith, A., Anand, H., Milosavljevic, S., Rentschler, K.M., Pocivavsek, A., Valafar, H. (2021). Application of machine learning to sleep stage classification. In 2021 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 349-354. https://doi.org/10.1109/CSCI54926.2021.00130

[24] Lee, Y.J., Lee, J.Y., Cho, J.H., Choi, J.H. (2022). Interrater reliability of sleep stage scoring: A meta-analysis. Journal of Clinical Sleep Medicine, 18(1): 193-202. https://doi.org/10.5664/jcsm.9538

[25] Perslev, M., Darkner, S., Kempfner, L., Nikolic, M., Jennum, P.J., Igel, C. (2021). U-Sleep: Resilient high-frequency sleep staging. NPJ Digital Medicine, 4(1): 72. https://doi.org/10.1038/s41746-021-00440-5

[26] Elreedy, D., Atiya, A.F., Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Machine Learning, 113(7): 4903-4923. https://doi.org/10.1007/s10994-022-06296-4