MSSC-BIMAMBA: MULTIMODAL SLEEP STAGE CLASSIFICATION WITH BIDIRECTIONAL MAMBA

Anonymous authors
Paper under double-blind review

ABSTRACT

Monitoring sleep states is essential for evaluating sleep quality and diagnosing sleep disorders. Traditional manual staging is time-consuming and prone to subjective bias, often resulting in inconsistent outcomes. Here, we developed an automated model for sleep staging to enhance diagnostic accuracy and efficiency. Considering the characteristics of polysomnography (PSG) multi-lead sleep monitoring, we designed a multimodal sleep state classification model, MSSC-BiMamba, that combines an Efficient Channel Attention (ECA) mechanism with a Bidirectional State Space Model (BSSM). The ECA module allows for weighting data from different sensor channels, thereby amplifying the influence of diverse sensor inputs. Additionally, the implementation of bidirectional Mamba (BiMamba) enables the model to effectively capture multi-dimensional features and long-range dependencies of PSG data. The developed model demonstrated impressive performance on sleep stage classification tasks on the ISRUC-S3 and ISRUC-S1 datasets, respectively, including healthy and unhealthy sleep patterns. Our model, which can effectively handle diverse sleep conditions, is the first to apply BiMamba to sleep staging with multimodal PSG data, showing substantial gains in computational and memory efficiency over traditional Transformer-style models. This method enhances sleep health management by making monitoring more accessible and extending advanced healthcare through innovative technology.

1 INTRODUCTION

Sleep, a fundamental pillar of human health and well-being, involves a complex series of stages in cognitive function and overall wellness Peter-Derex et al. (2015). However, modern lifestyles, work-related stress, and environmental factors have adversely impacted the sleep duration, continuity, and quality of many individuals, leading to a concerning rise in sleep problems, especially among children The Lancet Diabetes (2024). The accurate classification of sleep stages and assessment of sleep health are vital for the effective diagnosis and treatment of sleep disorders. Traditionally, polysomnography (PSG) has been considered the gold standard for monitoring sleep stages, but it requires specialized equipment and is typically confined to clinic settings. This limitation highlights the pressing need for more accessible and less intrusive approaches to sleep analysis.

Recent advances in machine learning, particularly deep learning, have paved the way for innovative approaches to sleep stage classification Loh et al. (2020). Among these, the Transformer model Vaswani et al. (2017), known for its effectiveness in handling sequential data primarily in the field of natural language processing, offers promising avenues for sleep data analysis due to its ability to capture long-range dependencies in time-series data. However, the Transformer's self-attention mechanism, which evaluates interactions across all sequence positions, significantly slows down as data length increases, and hence suffers from low efficiency due to heavy computational demands and limited scalability with large datasets. This necessitates more efficient algorithms or adaptations of existing models to optimize performance in tasks like sleep stage classification and anomaly detection without compromising computational resources.

Modern state space models (SSMs) particularly effective in capturing long-range dependencies, have evolved significantly with recent innovations. The Mamba model Gu & Dao (2023) advances SSMs by integrating time-varying parameters and a hardware-aware algorithm, enhancing training and inference efficiency dramatically. Despite its superior scalability and potential as an alternative to

Transformer in language modeling, Mamba is limited by its unidirectional approach and lack of positional awareness. To address these limitations, Vision Mamba Zhu et al. (2024) employs the Bidirectional State Space Model (BSSM) to leverage dynamics in both forward and reverse directions, significantly boosting prediction accuracy for sequential tasks. This bidirectional enhancement ensures high accuracy and enhances efficiency, making Vision Mamba ideal for large-scale studies and potentially suitable for real-time sleep monitoring applications.

In this study, we leverage the recent progress in deep learning by integrating the powerful capabilities of the Transformer architecture with the Bidirectional State Space Model (BSSM) from Vision Mamba. This novel combination harnesses the strengths of both approaches to effectively capture intricate temporal patterns within sleep data, enabling precise sleep stage classification. The bidirectional nature of the BSSM allows for comprehensive analysis of the temporal context, while the Transformer's deep learning capabilities facilitate the processing of large-scale, multimodal sleep monitoring datasets. By synergizing these cutting-edge techniques, our proposed model offers a robust and efficient solution for accurate sleep stage classification, pushing the boundaries of what is possible in the field of sleep medicine.

The main contributions of this paper are as follows:

- **Innovative Model Architecture**: We introduce the MSSC-BiMamba model with an innovative architecture, which combines the Efficient Channel Attention mechanism with BiMamba, tailored specifically for sleep stage classification.
- **Enhanced Efficiency**: The MSSC-BiMamba model, employing complex multimodal PSG data, significantly improves computational and memory efficiency, and bridges the gap between intricate clinical evaluations and practical, real-time monitoring solutions.
- **Superior Results**: Our model demonstrates superior performance on the ISRUC-S1 and ISRUC-S3 datasets, thereby confirming its generalizability and effectiveness across different sleep-related tasks.

2 RELATED WORK

2.1 CLASSIFICATION OF SLEEP STAGES

Traditionally, the classification of sleep stages is performed by experienced sleep specialists or physicians, who categorize 30-second or specific time intervals of polysomnographic (PSG) data into various sleep stages according to established sleep assessment criteria Liang et al. (2012). This process is known to be both laborious and time-consuming Malhotra et al. (2013). In contrast, machine learning algorithms may require significantly shorter durations to accomplish the same classification task Chang et al. (2019), while experts have traditionally extracted features from the time domain Sharma et al. (2017), frequency domain Zoubek et al. (2007), and time-frequency domain Al-Salman et al. (2019) to preprocess data for machine learning methods. The selection of effective features remains a critical issue in enhancing the classification performance of traditional classifiers.

The emergence of deep learning has further advanced the automation process, enabling direct extraction of complex features from raw data, thus reducing the preprocessing workload Sekkal et al. (2022). Therefore, an increasing number of deep learning methods are being applied to sleep stage classification tasks, including CNNs Phan et al. (2018), RNNs Zhu & Liang (2020), GCNs Jia et al. (2021), and others. However, CNNs may struggle to capture long-term dependencies in time-series data effectively. While RNNs are capable of handling sequential data and capturing long-term dependencies in time series, they are prone to issues such as vanishing or exploding gradients during training. Additionally, GCNs exhibit lower efficiency in processing large-scale graph data.

Ji et al. Ji et al. (2023; 2024) proposed several sleep stage classification models. These studies transformed PSG data into the frequency domain space, combined with time-domain signals as input to the model. They then utilized 3D CNNs, 2D CNNs, and GCN networks for classification, achieving satisfactory results. Compared to these methods, we propose an approach that solely utilizes time-domain signals but achieves superior performance.

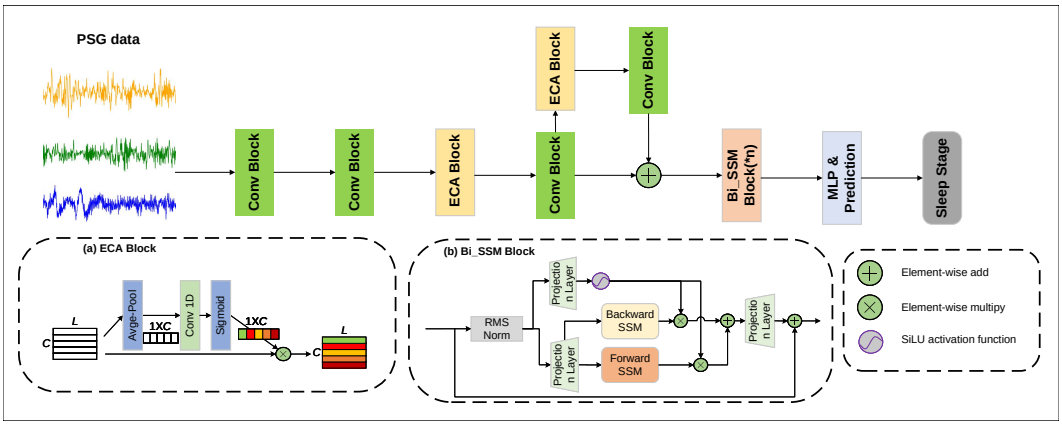

Figure 1: The architecture of the proposed MSSC-BiMamba model.

2.2 State space model and Mamba

Recent research advances have led to a surge of interest in state-space models (SSMs). Originating from the classical Kalman filtering model Kalman (1960), SSMs excel at capturing long-term dependencies. Researchers have proposed several SSM-based methods, such as Structured State Space Sequence Models (S4) Gu et al. (2021) and S4D Smith et al. (2022), for handling sequential data from various tasks and modalities, particularly in modeling remote dependencies. Due to their convolutional and near-linear computations, they exhibit high efficiency in processing long sequences. In recent studies, Mamba Gu & Dao (2023) integrated time-varying parameters into SSMs and proposed a hardware-aware algorithm to achieve highly efficient training and inference. Compared to the original Mamba, the Bidirectional Mamba offers higher efficiency and performance. Based on this structure, we explore the temporal PSG signals.

3 METHOD

3.1 Overview of the framework

The proposed MSSC-BiMamba model, depicted in Figure 1, addresses a critical challenge in sleep medicine: the classification of sleep stages utilizing PSG data, such as EEG and EOG signals. This innovative model is specifically designed to enhance diagnostic accuracy and efficiency, thereby boosting a deeper understanding of sleep health in clinical environments, and facilitating the early diagnosis of sleep disorders.

The architecture of MSSC-BiMamba incorporates the Efficient Channel Attention (ECA) module to focus on salient features in the data effectively, followed by the Bidirectional Mamba (BiMamba) module, which processes these features to capture complex temporal relationships. This setup ensures a robust analysis of sleep patterns. In the experimental setup, we detail the configuration of hyperparameters and the specific conditions under which the model operates. We conclude by describing the evaluation metrics used to assess the performance of the model, ensuring a comprehensive understanding of its effectiveness in clinical applications.

3.2 The Efficient Channel Attention module

In recent years, the introduction of channel attention mechanisms into convolutional blocks has garnered widespread attention. A prominent method in this domain is SENet (Squeeze-and-Excitation Networks)Hu et al. (2018), which learns channel attention for each convolutional block and significantly enhances the performance of various deep CNN architectures. While these methods achieve high accuracy, they often lead to increased model complexity and substantial computational burden. To address these challenges, Wang et al.Wang et al. (2020) proposed the Efficient Channel Attention (ECA) module. The ECA module avoids dimensionality reduction and effectively captures

cross-channel interactions, ensuring both efficiency and effectiveness. This module has achieved commendable results in the field of image processing.

The channel attention mechanism is critical for dynamically adjusting the channel responses of feature maps by learning the importance of each channel. This process enhances the representational capacity of neural networks by allowing for the dynamic recalibration of feature channel correlations. Specifically, for time series data, this involves adjusting multiple PSG channel data to capture temporal patterns and dependencies more effectively. Building on the ECA module, we propose an adaptation tailored to time series data (Figure 1 .a). Our approach ensures the suitability and efficacy of the ECA mechanism in handling the unique characteristics of time series, such as temporal dependencies and sequential patterns.

Let x_{ci} denote the input feature map, where *c* is the number of channels and *L* represents the length of the feature map. First, the global spatial information of each channel is computed through a global average pooling (GAP) operation, resulting in channel descriptors S_c.

$$S_c = \frac{1}{L} \sum_{i=1}^{L} x_{ci} \tag{1}$$

Subsequently, channel weights W are obtained through convolutional layers and activation functions.

$$W = \sigma(Conv1d(S)) \tag{2}$$

Finally, the learned channel weights are applied to the original feature map, yielding the weighted feature map $\tilde{X}_{ci}$.

$$\tilde{X}_{ci} = W_c \cdot X_{ci} \tag{3}$$

In this way, the network can strengthen the important channel features for sleep stage classification tasks while suppressing those unimportant channel features, thereby enhancing the model's ability to classify sleep stages.

3.3 BIDIRECTIONAL MAMBA

The S4 and Mamba structures discretize the state space representation of continuous systems. They utilize zero-order hold (ZOH) to maintain the dynamic characteristics of the system, addressing the limitation of direct implementation of continuous systems on digital computers, as digital computers can only process discrete signals. Through precise sampling and holding processes, stable and efficient digital implementation of continuous systems is permitted. The continuous state *h*(*t*) and input *x*(*t*) can be represented as:

$$h'(t) = Ah(t) + Bx(t) \tag{4}$$
$$y(t) = Ch(t) \tag{5}$$

where $A \in \mathbb{R}^{N \times N}$ is the state transition matrix, and $B \in \mathbb{R}^{N \times 1}$, $C \in \mathbb{R}^{N \times 1}$ is the input and output matrix.

Through the time scale parameter Δ, the continuous parameters *A* and *B* are transformed into discrete parameters, resulting in $\bar{A}$, $\bar{B}$. Zero-order hold (ZOH) is a method of maintaining the value of a signal unchanged during sampling and holding until the next sampling point. After discretizing *A* and *B*, the discrete versions using the time scale parameter Δ can be rewritten in the following form:

$$\begin{aligned} h_t &= \bar{A}h_{t-1} + \bar{B}x_t \\ y_t &= Ch_t \end{aligned} \tag{6}$$

where, $\bar{A} = \exp(\Delta A)$, $\bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \cdot B$, finally, the model's output is obtained through a global convolution.

$$\begin{gathered} \bar{K} = (C\bar{B}, C\overline{AB}, \cdots, C\bar{A}^{L-1}\bar{B}), \\ y = x \cdot \bar{K} \end{gathered} \tag{7}$$

where $\bar{K} \in \mathbb{R}^L$ is the structured convolutional kernel, and L is the length of the input sequence x.

As shown in Figure 1 .b, the bidirectional Mamba utilizes both forward and backward modules to expedite the SSM process and enhance the capability of acquiring contextual information. Specifically, it scans the sequence once from start to end and again from end to start. Subsequently, it averages the outputs from both scans to obtain a comprehensive representation.

The BSSM Block operates on an input $x_t \in \mathbb{R}^{B \times L \times D}$, where B is the batch size, L is sequence length , and D is hidden dimension at each time step t. Initially, it performs linear projections on x_t , expanding the hidden dimension to D , resulting in x and z (Algorithm.1,2). These projections undergo further processing through 1D convolution and SiLu activation (Algorithm.5). The block's core features a discretized BiSSM with parameters adjusted based on the input (Algorithm.6-13). This discretized BiSSM, along with x, generates the state representation $\{y_i | i \in \{\text{forward}, \text{backward}\}\}$. After y_i is combined with a residual connection from z after applying SiLu activation, a linear projection then delivers the output y_t at time step t (Algorithm.16). Overall, the BSSM Block effectively processes sequential information by adapting to input variations and integrating BiSSM capabilities.

Algorithm 1 BSSM Block Process Algorith

Require: $x_t : (B, L, D)$
Ensure: $y_t : (B, L, D)$
1: $x : (B, L, E) \leftarrow \text{Linear}(x_t)$
2: $z : (B, L, E) \leftarrow \text{Linear}(x_t)$
3: **process Bidirectional Mamba**
4: **for** each direction $i \in \{\text{forward}, \text{backward}\}$ **do**
5: $x' : (E, L, ED) \leftarrow \text{SiLU}(\text{Conv1d}_i(x))$
6: $B_i : (B, L, N) \leftarrow \text{Linear}(x')$
7: $C_i : (B, L, N) \leftarrow \text{Linear}(x')$
8: **softplus ensures positive** Δ_i
9: $\Delta_i : (B, L, D) \leftarrow \log(1 + \exp(\text{Linear}(x') + \text{Parameter}^{\Delta_i}))$
10: **shape of Parameter**$_i^A$ **is** (E, N)
11: $A_i : (B, L, D, N) \leftarrow \Delta_i \otimes \text{Parameter}_i^A$
12: $B_i : (B, L, D, N) \leftarrow \Delta_i \otimes B_i$
13: $y_i : (B, L, ED) \leftarrow \text{SSM}(A_i, B_i, C_i)(x')$
14: **end for**
15: **residual connection**
16: $y_t : (B, L, D) \leftarrow \text{Linear}((\sum_{i \in \{\text{forward,backward}\}}) y_i \odot \text{SiLU}(z))$
17: **return** y_t

4 EXPERIMENTS

4.1 DATASETS

The ISRUC-Sleep dataset Khalighi et al. (2016) comprises data from both healthy subjects and individuals with sleep disorders. Specifically, the ISRUC-S3 subset contains data from 10 healthy subjects, while the S1 subset includes data from 100 participants with sleep disorders. The data provider performed preprocessing on all PSG recordings, which involved the following steps: 1) 50 Hz power-line noise removal using a notch filter; 2) for the electroencephalogram (EEG) and electrooculogram (EOG) data, a bandpass Butterworth filter was applied to obtain waves in the frequency range of 0.3 Hz to 35 Hz; 3) the electromyogram (EMG) data were filtered using a low cutoff frequency of 10 Hz and a high cutoff frequency of 70 Hz. To mitigate the impact of noise,

the last 30 epochs from each subject were removed. This approach aimed to eliminate potentially unreliable or contaminated data points that could adversely affect the accuracy and robustness of the subsequent analyses and sleep stage classification.

Table 1: Detailed information of the ISRUC-Sleep datasets used for sleep stage classification.

Subset	Subject Number	W	N1	N2	N3	REM	Total
ISRUC-S3	10	1674	1217	2616	2016	1066	8589
ISRUC-S1 (50)	50	10097	5555	13250	8675	5779	43356
ISRUC-S1 (all)	100	20098	11062	27511	17251	11265	87187

4.2 Experimental setup

According to the American Academy of Sleep Medicine (AASM) guidelines Berry et al. (2012), the sleep-wake cycle can be classified into five stages: wakefulness (W), rapid eye movement (REM) sleep, and three non-rapid eye movement (NREM) stages, namely N1, N2, and N3. The NREM stages are further categorized based on the depth of sleep, with N1 representing the lightest stage of sleep, N2 being a deeper stage, and N3 corresponding to the deepest stage of sleep, also known as slow-wave sleep.

For the classification of sleep stages, we utilized all data from the S3 dataset and 50 participants with odd-numbered IDs from the S1 dataset (Table 1), and selected PSG data from ten channels, including EEG (F3-A2,C3-A2,O1-A2, F4-A1,C4-A1,O2-A1), EOG (LOC-A2,ROC-A1), ECG(X2) and Chin EEG (X1).

To prepare the data for analysis, we first applied a downsampling technique, reducing the sampling rate from 200 Hz to 100 Hz. This step helps to minimize computational complexity while preserving the essential information in the signal. Subsequently, we performed slicing operations on the data from each channel. The data input into the network comprised 1000 points (30 s), about 33.33 Hz. The ISRUC-S3 dataset employed a 10-fold cross-validation, while the ISRUC-S1 dataset utilized a 25-fold cross-validation. Cross-validation was independently performed for each subject to ensure that data from the same subject did not appear simultaneously in both the training and validation sets. The experimental setup and hyperparameters are summarized in Appendix A.

5 Experiments and numerical analysis

5.1 Classification experiments on the ISRUC-S3 and ISRUC-S1 datasets

To assess the performance of our model, we conducted experiments using four different combinations on the ISRUC-S3 dataset, testing the functionality of different modules. The experimental results (Table 2) indicate that incorporating the ECA module can notably enhance performance. Compared to the model without the ECA module, performance improvements are observed across all metrics except for F1 in the W stage, where the performance remains relatively consistent. To evaluate the performance of models with different depths, we configured BiMamba modules with 1 layer, 2 layers, 3 layers, and 10 layers. Through comparison, we found that increasing the depth of the model does not necessarily lead to performance improvement. However, for the challenging N1 stage with a limited sample size, some improvement is observed. The model with CNN, ECA, and 1-layer BiMamba performed best. The confusion matrices are presented in Appendix B.

When compared to previous models (Table 2), our model outperforms them in most performance metrics on the ISRUC-S3 dataset. However, it exhibits marginally lower F1-scores in the Wake (W) and Rapid Eye Movement (REM) stages compared to the JK-STGCN Ji et al. (2022) and MixSleepNet Ji et al. (2024) models. Specific comparisons of the F1 scores for each sleep stage are provided in Appendix B. Additionally, our model benefits from a significantly reduced parameter count, making it more efficient than earlier deep-learning models in terms of computational resources and memory usage. This efficiency makes our model particularly suitable for deployment in clinical settings where computational power may be limited.

Table 2: Comparison with other methods on the ISRUC-S3 dataset.

Model	Parameter	ACC	F1	Kappa
RF Memar & Faradji (2017)	<0.1M	0.702	0.685	0.616
DeepSleepNet Supratak et al. (2017)	21M	0.719	0.696	0.643
GraphSleepNet Jia et al. (2020)	-	0.786	0.770	0.724
JK-STGCN Ji et al. (2022)	-	0.831	0.814	0.782
MixSleepNet Ji et al. (2024)	2.4M	0.830	0.821	0.782
CNN+1BiMamba	0.47M	0.845	0.817	0.794
MSSC-BiMamba	**0.47M**	**0.852**	**0.824**	**0.803**
CNN+ECA+2BiMamba	0.73M	0.847	0.819	0.796
CNN+ECA+3BiMamba	0.99M	0.850	0.822	0.800
CNN+ECA+10BiMamba	2.79M	0.850	0.823	0.800

After conducting a comparative analysis on the ISRUC-S3 dataset, we finalized the model parameters and selected the MSSC-BiMamba (CNN+ECA+1BiMamba) model. Subsequently, we performed experiments on the ISRUC-S1 dataset. As illustrated in Table 3, our model outperforms other models in most performance metrics. Although it exhibits slightly lower F1-scores in the W and N1 stages compared to Mixsleepnet and JK-STGCN, it excels in other stages, highlighting its overall effectiveness. Given that the model parameters were optimized for ISRUC-S3, the performance on ISRUC-S1 could likely be further enhanced with additional tuning specific to this dataset. Nonetheless, the current results underscore the potential and adaptability of the model in diverse sleep stage classification tasks.

Table 3: Comparison with other methods on the ISRUC-S1 dataset.

Method	ACC	F1	Kappa
RF Memar & Faradji (2017)	0.699	0.649	0.607
DeepSleepNetSupratak et al. (2017)	0.730	0.691	0.654
GraphSleepNetJia et al. (2020)	0.780	0.751	0.715
JK-STGCNJi et al. (2022)	0.820	0.798	0.767
MixsleepnetJi et al. (2024)	0.813	0.787	0.757
MSSC-BiMamba	**0.830**	**0.801**	**0.773**

To provide a more intuitive comparison of the model's performance, the sleep state classifications for participants are presented, showing the alignment between expert-labeled outcomes and those produced by our models. Although minor discrepancies occur during transitions between sleep stages, the majority of the classifications closely match expert assessments, as illustrated in Appendix C.

Therefore, the developed model demonstrated impressive performance on sleep stage classification tasks on both the ISRUC-S3 and ISRUC-S1 datasets. These datasets respectively contain clinical PSG data from populations with healthy and unhealthy sleep patterns. This distinction highlights the model's capability to effectively differentiate and analyze sleep stages across diverse health conditions, showcasing its potential for broad applications in sleep medicine. Especially, our efficient model can free clinical physicians from the tedious task of manual staging, allowing them to focus more on supervising and correcting the automated classifications, as well as devoting more energy to the treatment and care of patients. This automated classification method not only enhances diagnostic accuracy, but also significantly improves efficiency when dealing with large volumes of data, bringing significant technological innovations to the field of sleep medicine.

5.2 Cross-validation experiments

In order to assess the generalization ability of the model across different datasets, we performed cross-validation experiments using models trained on ISRUC-S1(50) and ISRUC-S3 datasets, utilizing ISRUC-S1(50), ISRUC-S1(100), and the entire ISRUC-S3 dataset as test sets to evaluate the performance of all models through K-fold validation and finally computed the average performance.

Table 4: Comparison of cross-experiment results between the ISRUC-S3 and ISRUC-S1 models.

Model	Datasets	ACC	F1	Kappa
ISRUC-S3	ISRUC-S3	0.857	0.851	0.817
	ISRUC-S1(50)	0.330	0.281	0.160
ISRUC-S1(50)	ISRUC-S1(50)	0.852	0.833	0.809
	ISRUC-S1(100)	0.808	0.788	0.752
	ISRUC-S3	0.797	0.774	0.737

From Table 4, it can be observed that the S1(50) model performs well on all three subsets, S1(50), S1(100), and S3, demonstrating its strong generalization ability. However, models trained on the S3 dataset do not perform as well on the S1 dataset, possibly due to differences between the S3 and S1 datasets, as well as the smaller size of the S3 dataset. Overall, the model's robust performance across various metrics underscores its potential for reliable and accurate sleep stage classification.

6 CONCLUSIONS

In this study, we have demonstrated the exceptional generalization capabilities and computational efficiency of the Mamba-driven deep learning framework in handling complex polysomnography (PSG) data, underscoring its potential to revolutionize sleep medicine. By assisting doctors in annotation, this innovative approach can transform medical experts from front-end operators to back-end supervisors, streamlining diagnostic workflows and optimizing resource utilization. Future research will focus on detailed analysis and model training for different sleep disorders. By incorporating rich features and optimizing algorithms, we aim to enhance the model's predictive accuracy and enable precise, real-time sleep stage recognition. Developing accurate, reliable, and efficient tools for sleep stage recognition and sleep disorder diagnosis will not only enhance patient care but also deepen our understanding of the intricate mechanisms underlying sleep health. This will provide valuable insights and support for the diagnosis and management of sleep disorders, ultimately improving patient care and outcomes. We remain committed to exploring the potential of the Mamba framework, striving for breakthroughs in sleep medicine. Our ongoing efforts will continue to push boundaries, driving progress in this field and paving the way for a brighter future.

REFERENCES

Wessam Al-Salman, Yan Li, and Peng Wen. Detection of eeg k-complexes using fractal dimension of time frequency images technique coupled with undirected graph features. *Frontiers in Neuroinformatics*, 13:45, 2019.

Richard B Berry, Rita Brooks, Charlene E Gamaldo, Susan M Harding, Carole Marcus, Bradley V Vaughn, et al. The aasm manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 176(2012): 7, 2012.

Shang-Yuan Chang, Bing-Chen Wu, Yi-Long Liou, Rui-Xuan Zheng, Pei-Lin Lee, Tzi-Dar Chiueh, and Tsung-Te Liu. An ultra-low-power dual-mode automatic sleep staging processor using neural-network-based decision tree. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 66 (9):3504–3516, 2019.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

Xiaopeng Ji, Yan Li, and Peng Wen. Jumping knowledge based spatial-temporal graph convolutional networks for automatic sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:1464–1472, 2022.

Xiaopeng Ji, Yan Li, and Peng Wen. 3dsleepnet: A multi-channel bio-signal based sleep stages classification method using deep learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:3513–3523, 2023. doi: 10.1109/TNSRE.2023.3309542.

Xiaopeng Ji, Yan Li, Peng Wen, Prabal Barua, and U Rajendra Acharya. Mixsleepnet: A multi-type convolution combined sleep stage classification model. *Computer Methods and Programs in Biomedicine*, 244:107992, 2024. ISSN 0169-2607.

Ziyu Jia, Youfang Lin, Jing Wang, Ronghao Zhou, Xiaojun Ning, Yuanlai He, and Yaoshuai Zhao. Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification. In *IJCAI*, volume 2021, pp. 1324–1330, 2020.

Ziyu Jia, Youfang Lin, Jing Wang, Xiaojun Ning, Yuanlai He, Ronghao Zhou, Yuhan Zhou, and H Lehman Li-wei. 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, 2021.

Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960.

Sirvan Khalighi, Teresa Sousa, José Moutinho Santos, and Urbano Nunes. Isruc-sleep: A comprehensive public dataset for sleep researchers. *Computer methods and programs in biomedicine*, 124:180–192, 2016.

Sheng-Fu Liang, Chin-En Kuo, Yu-Han Hu, and Yu-Shian Cheng. A rule-based automatic sleep staging method. *Journal of neuroscience methods*, 205(1):169–176, 2012.

Hui Wen Loh, Chui Ping Ooi, Jahmunah Vicnesh, Shu Lih Oh, Oliver Faust, Arkadiusz Gertych, and U Rajendra Acharya. Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020). *Applied Sciences*, 10(24):8963, 2020.

Atul Malhotra, Magdy Younes, Samuel T Kuna, Ruth Benca, Clete A Kushida, James Walsh, Alexandra Hanlon, Bethany Staley, Allan I Pack, and Grace W Pien. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep*, 36(4): 573–582, 2013.

Pejman Memar and Farhad Faradji. A novel multi-class eeg-based sleep stage classification system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(1):84–95, 2017.

Laure Peter-Derex, Pierre Yammine, Hélène Bastuji, and Bernard Croisile. Sleep and alzheimer's disease. *Sleep medicine reviews*, 19:29–38, 2015.

Huy Phan, Fernando Andreotti, Navin Cooray, Oliver Y Chén, and Maarten De Vos. Joint classification and prediction cnn framework for automatic sleep stage classification. *IEEE Transactions on Biomedical Engineering*, 66(5):1285–1296, 2018.

Rym Nihel Sekkal, Fethi Bereksi-Reguig, Daniel Ruiz-Fernandez, Nabil Dib, and Samira Sekkal. Automatic sleep stage classification: From classical machine learning methods to deep learning. *Biomedical Signal Processing and Control*, 77:103751, 2022.

Rajeev Sharma, Ram Bilas Pachori, and Abhay Upadhyay. Automatic sleep stages classification based on iterative filtering of electroencephalogram signals. *Neural Computing and Applications*, 28:2959–2978, 2017.

Jimmy TH Smith, Andrew Warrington, and Scott W Linderman. Simplified state space layers for sequence modeling. *arXiv preprint arXiv:2208.04933*, 2022.

Akara Supratak, Hao Dong, Chao Wu, and Yike Guo. Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):1998–2008, 2017.

Endocrinology The Lancet Diabetes. Sleep: a neglected public health issue. *The Lancet Diabetes & Endocrinology*, 12(6):365, June 2024. ISSN 2213-8587, 2213-8595. doi: 10.1016/S2213-8587(24)00132-3.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11534–11542, 2020.

Fangqi Zhu and Qilian Liang. Ocrnn: An orthogonal constrained recurrent neural network for sleep analysis based on eeg data. *Ad Hoc Networks*, 104:102178, 2020.

Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024.

Lukáš Zoubek, Sylvie Charbonnier, Suzanne Lesecq, Alain Buguet, and Florian Chapotot. Feature selection for sleep/wake stages classification using data driven methods. *Biomedical Signal Processing and Control*, 2(3):171–179, 2007.

A EXPERIMENT DESIGN

Table 5: Experimental environment and hyperparameter descriptions.

Parameter	**Value**
GPU	NVIDIA GeForce RTX 4090
CPU	AMD Ryzen 9 7900 12-Core Processor
Pytorch	Torch 2.1.1+cu118
Python	3.10.13
Epoch	40
Batch Size	100
Learning Rate	0.001
Weight Decay	0.0001
Dropout	0.2
Optimizer	Adam

B ADDITIONAL RESULTS

B.1 CONFUSION MATRIX FOR ABLATION EXPERIMENTS

The confusion matrices in Figure 2 illustrate the performance of different models on the ISRUC-S3 dataset. By comparing panels (a) and (b), it is evident that the inclusion of the Efficient Channel Attention (ECA) module enhances classification performance for all stages except N1. Further comparison of panels (b), (c), and (d) reveals that deepening the network improves the recognition performance for the N1 stage. However, this improvement comes at the cost of reduced performance for other stages.

Figure 2: The confusion matrices for experiments on the ISRUC-S3 dataset: a, CNN+1bimamba; b, (MSSC-BiMamba) CNN+ECA+1bimamba; c, CNN+ECA+2bimamba; d, CNN+ECA+10bimamba.

B.2 Comparison of F1 scores for each sleep stage classification

Table 6: Comparison F1 Score all class with other methods on the ISRUC-S3 dataset.

Method	W	N1	N2	N3	REM
RF Memar & Faradji (2017)	0.838	0.470	0.671	0.763	0.684
DeepSleepNet Supratak et al. (2017)	0.831	0.463	0.742	0.851	0.595
GraphSleepNet Jia et al. (2020)	0.864	0.540	0.782	0.869	0.793
JK-STGCN Ji et al. (2022)	**0.900**	0.598	0.826	0.901	0.845
MixSleepNet Ji et al. (2024)	0.899	0.625	0.819	0.899	**0.860**
CNN+1BiMamba	0.871	0.620	0.834	0.910	0.852
MSSC-BiMamba	0.886	0.624	**0.841**	0.915	0.854
CNN+ECA+2BiMamba	0.878	0.625	0.831	0.914	0.847
CNN+ECA+3BiMamba	0.871	0.629	0.840	0.915	0.855
CNN+ECA+10BiMamba	0.879	**0.635**	0.838	**0.916**	0.846

Table 7: Comparison F1 Score all class with other methods on the ISRUC-S1 dataset.

Method	W	N1	N2	N3	REM
RF Memar & Faradji (2017)	0.841	0.307	0.705	0.750	0.640
DeepSleepNetSupratak et al. (2017)	0.850	0.385	0.739	0.830	0.648
GraphSleepNetJia et al. (2020)	0.889	0.463	0.763	0.825	0.813
JK-STGCNJi et al. (2022)	0.895	**0.550**	0.811	0.883	0.850
MixsleepnetJi et al. (2024)	**0.908**	0.512	0.799	0.871	0.844
MSSC-BiMamba	0.901	0.547	**0.812**	**0.885**	**0.860**

C MSSC-BIMAMBA PREDICTIONS COMPARED TO EXPERTS

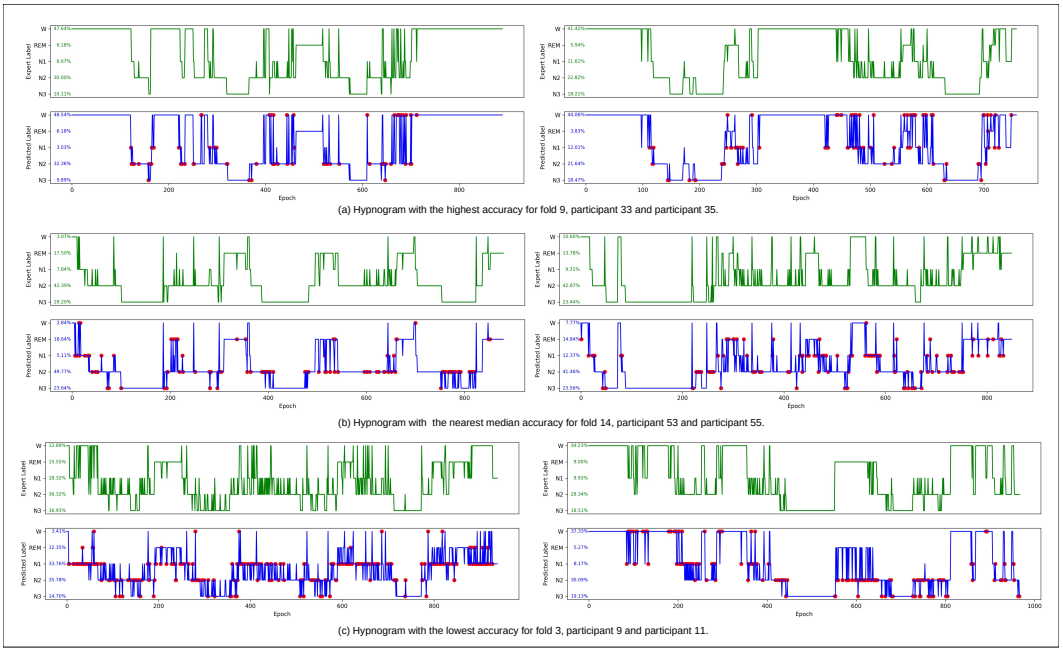

Figure 3: Visualization of the highest, nearest median, and lowest scoring in the ISRUC-S1 dataset between expert-labeled outcomes and predictions by our ISRUC-S1 model

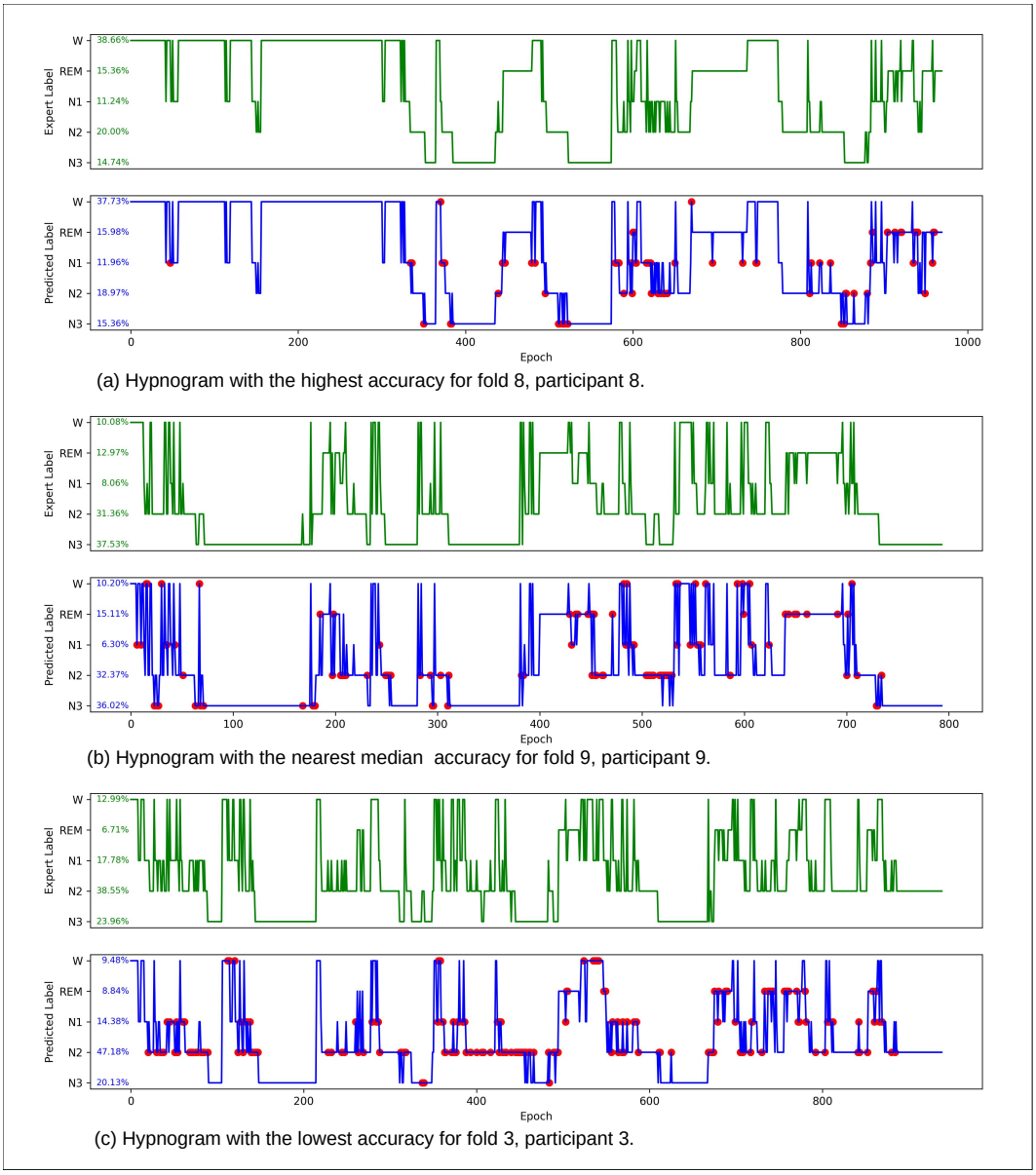

Figure 4: Visualization of the highest, nearest median, and lowest scoring in the ISRUC-S3 dataset between expert-labeled outcomes and predictions by our ISRUC-S3 model