EC-SLEEPMAMBA: AN EEG-GUIDED CROSS-MODAL MAMBA FRAMEWORK FOR SLEEP STAGING

Anonymous authors
Paper under double-blind review

ABSTRACT

Automatic sleep staging is an important component of quantitative sleep assessment and provides fundamental support for sleep analysis. Despite recent advances, existing multimodal fusion-based approaches face two challenges: (i) The symmetric fusion strategy ignores the asymmetric clinical prior that EEG is the primary modality, while EOG and EMG serve as auxiliary signals, overlooking the inherent discrepancy in their information contributions. (ii) Jointly modeling intra-epoch dynamics with inter-epoch stage transitions across long-duration physiological signals remains computationally demanding. To address these issues, we propose EC-SleepMamba, an EEG-guided cross-modal Mamba framework that leverages modality complementarity and temporal context to improve staging robustness. Specifically, modality-specific dual-scale CNNs are first introduced to extract high-resolution local features. During sequence encoding, an EEG-guided interaction gate is proposed to condition the auxiliary Cross-Mamba branches on EEG-derived context, enabling directional primary-to-auxiliary information regulation during representation encoding. Finally, a global bi-Mamba module is adopted to capture dynamic transition patterns across epochs. Experiments on Sleep-EDF-39 and Sleep-EDF-153 show that EC-SleepMamba outperforms five competing methods and improves the discrimination of highly confusable stages, demonstrating its effectiveness for sleep staging.

1 INTRODUCTION

Automatic sleep staging (ASS) is essential for assessing sleep quality and diagnosing sleep disorders, and it has substantial value for daily sleep health monitoring (Ohayon, 2011; Institute of Medicine, 2006). In clinical practice, experts divide whole-night polysomnography (PSG) recordings into 30-s epochs and assign each epoch to one of five sleep stages according to the AASM rules (Berry et al., 2012; Iber et al., 2007). Because manual scoring is labor-intensive and subject to inter-scorer variability (Rosenberg & Van Hout, 2013), numerous machine-learning methods have been developed for ASS (Fiorillo et al., 2019; Ronzhina et al., 2012). Deep models, particularly CNNs, have further advanced the automatic extraction of discriminative representations from raw physiological signals (Supratak et al., 2017; Tsinalis et al., 2016). Despite this progress, two challenges remain open for multimodal sleep staging.

1) The asymmetric clinical roles of multimodal signals should be explicitly modeled. Clinical sleep scoring typically does not rely equally on EEG, EOG, and EMG. EEG is widely considered to directly reflect brain activity and serves as the primary source of evidence for the underlying sleep state. Specifically, N2 and N3 are primarily identified based on EEG events such as sleep spindles, K-complexes, and slow-wave activity. By contrast, while the EEG patterns for N1 and REM can be ambiguous, slow and rapid eye movements in EOG provide crucial discriminative cues; additionally, reduced submental EMG tone offers vital auxiliary evidence for REM (Berry et al., 2012; Iber et al., 2007; Keenan & Hirshkowitz, 2011; Jia et al., 2021). This establishes an asymmetric clinical logic where EEG is the primary modality, and EOG/EMG are auxiliary. Existing multimodal methods typically exploit complementary signals through multichannel convolution, post-encoding feature fusion, or cross-modal attention among modality-specific representations (Chambon et al., 2018; Andreotti et al., 2018; Jia et al., 2020; 2021; Wang et al., 2025). Attention-based methods can even assign stage-adaptive importance to encoded features (Jia et al., 2021; Wang et al., 2025). However, these cross-modal interactions still largely treat modality streams as parallel information

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

sources. They fail to explicitly model the asymmetric clinical prior of "EEG as primary, EOG/EMG as auxiliary" and make limited use of EEG-derived brain-state context to regulate the propagation of EOG/EMG information during the formation of auxiliary sequence representations.

2) Intra-epoch dynamics and inter-epoch stage transitions should be modeled jointly and efficiently. Many existing works utilize recurrent neural networks (RNNs) and their variants to model temporal dependencies through hidden-state propagation (Phan et al., 2019; 2022a). However, RNN-based architectures face serious optimization difficulties and limited parallel efficiency on extremely long sequences (Bai et al., 2018). Recently, Transformers have gradually become an alternative to RNNs in the ASS field because of their powerful global-context modeling capabilities (Phan et al., 2022b). Nevertheless, standard Transformers model temporal interactions through self-attention rather than explicitly maintaining a recurrent hidden state. For sleep staging, we argue that recurrent state propagation is well aligned with the temporal structure of the task: it is suitable for representing both the ordered physiological evolution within each epoch and the structured sleep-stage transitions across adjacent epochs (Phan et al., 2019; 2022a; Bulatov et al., 2022; Zhou et al., 2023; Shen et al., 2024). Furthermore, the quadratic complexity of self-attention introduces substantial computational and memory overhead when processing high-sampling-rate continuous PSG sequences (Vaswani et al., 2017). Meanwhile, many existing temporal models primarily emphasize either intra-epoch waveform dynamics or inter-epoch sleep-stage transitions (Shen et al., 2024). Therefore, efficiently modeling fine-grained intra-epoch physiological dynamics and inter-epoch sleep-stage transitions through a unified backbone that combines recurrent state updates, linear scaling, and efficient parallel computation remains underexplored.

To address these challenges, we propose EC-SleepMamba, an EEG-guided cross-modal Mamba framework designed to exploit modality complementarity and temporal context for robust sleep staging. First, modality-specific Dual-Scale CNNs are proposed to extract high-resolution local features that preserve both coarse physiological rhythms and transient sleep-related waveforms. During sequence encoding, we introduce a Local Bi-Mamba to capture bidirectional intra-epoch EEG dynamics and propose an EEG-guided Interaction Gate that leverages EEG-derived sleep-stage semantics to modulate the internal selective Z-gates of the EOG and EMG Cross-Mamba branches, thereby enabling adaptive and effective information exchange across the three modalities. This primary-to-auxiliary interaction regulates auxiliary information flow before the corresponding sequence representations are finalized, rather than relying only on post-encoding fusion. Finally, Global Bi-Mamba captures inter-epoch stage transitions. Through this hierarchical design, EC-SleepMamba aligns cross-modal interaction with the asymmetric clinical roles of EEG, EOG, and EMG, while jointly modeling intra-epoch dynamics and inter-epoch stage transitions within a unified state-space backbone.

The main contributions are summarized as follows:

- We propose an EEG-guided cross-modal Mamba framework that incorporates asymmetric clinical priors into cross-modal interaction, and unifies intra-epoch dynamics with inter-epoch contextual dependencies in a shared state space, thereby enabling more discriminative multimodal representations for sleep staging.
- We present an EEG-guided Interaction Gate that injects EEG-derived sleep-stage semantics into the internal selective Z-gates of the EOG and EMG Cross-Mamba branches. This directional primary-to-auxiliary regulation enables cross-modal adaptation during auxiliary representation formation rather than relying solely on post-encoding fusion.
- We develop a hierarchical state-space backbone in which Local Bi-Mamba and Cross-Mamba capture intra-epoch physiological dynamics, while Global Bi-Mamba models inter-epoch sleep-stage transitions with linear computational complexity in sequence length.

2 Related Work

Multimodal Learning for Sleep Staging Multimodal ASS exploits complementary physiological signals collected by PSG. Existing studies differ mainly in where multimodal information is combined. Chambon et al. learn linear spatial filters and use partially separate EEG/EOG and EMG pipelines before a common classifier, while Andreotti et al. evaluate multichannel CNNs and demonstrate the benefit of adding EOG and EMG to EEG (Chambon et al., 2018; Andreotti et al.,

2018). These architectures establish multimodal complementarity, but do not introduce an explicit stage-dependent modality-weighting mechanism. SleepPrintNet constructs independent EEG temporal, EEG spectral-spatial, EOG, and EMG feature extractors, and fuses their outputs after branch-wise encoding (Jia et al., 2020). SalientSleepNet introduces multimodal attention to strengthen stage-relevant EEG/EOG features, whereas SleepWaveNet combines subject-adaptive wave extraction with multimodal attention (Jia et al., 2021; Wang et al., 2025). Accordingly, adaptive modality emphasis is specific to the latter attention-based models rather than a property shared by all multimodal methods. Even for these models, adaptive emphasis and directional conditioning are different: attention aggregates or reweights encoded features, but does not explicitly use EEG context to regulate which EOG/EMG events are retained during auxiliary sequence encoding. EC-SleepMamba addresses this gap by placing directional EEG-to-auxiliary regulation inside the EOG/EMG Cross-Mamba encoders.

Temporal Context Learning for Sleep Staging Temporal modeling in ASS has progressed from isolated-epoch waveform extraction to sequence-level contextual reasoning. CNN-based methods such as DeepSleepNet and TinySleepNet learn discriminative local patterns from raw EEG; DeepSleepNet further combines these features with recurrent transition modeling, whereas U-Time formulates sleep staging as dense temporal segmentation (Supratak et al., 2017; Supratak & Guo, 2020; Perslev et al., 2019). To exploit the continuity and constrained transitions of sleep stages, SeqSleepNet introduces hierarchical sequence-to-sequence learning, while XSleepNet combines multiple recurrent views for contextual staging (Phan et al., 2019; 2022a). SleepTransformer replaces recurrent context modeling with self-attention to capture non-local dependencies (Phan et al., 2022b). In parallel, structured state-space models such as S4 and Mamba provide recurrent state updates, parallel training, and linear sequence scaling (Gu et al., 2022; Gu & Dao, 2024). BiT-MamSleep applies bidirectional Mamba to single-channel EEG staging, demonstrating the potential of state-space models for sleep sequence modeling (Zhou et al., 2024). However, it does not address multimodal asymmetry or integrate token-level cross-modal interaction with epoch-level stage-transition modeling. Consequently, hierarchical temporal modeling and asymmetric cross-modal regulation remain insufficiently unified in a multimodal state-space framework. EC-SleepMamba applies Mamba at both token and epoch levels and further conditions auxiliary token encoding on EEG context, thereby unifying hierarchical temporal modeling with asymmetric cross-modal regulation.

3 EEG-Guided Cross-Modal Mamba Framework

A whole-night polysomnography (PSG) recording is represented as a sequence of consecutive sleep epochs. For the i-th epoch, the multimodal input is denoted as $\mathbf{x}_i = \{\mathbf{x}_i^E, \mathbf{x}_i^O, \mathbf{x}_i^M\}$, where $\mathbf{x}_i^E \in \mathbb{R}^{C_E \times T}$, $\mathbf{x}_i^O \in \mathbb{R}^{C_O \times T}$, and $\mathbf{x}_i^M \in \mathbb{R}^{C_M \times T}$ represent EEG, EOG, and EMG signals, respectively. In this work, $C_E = 2$, $C_O = 1$, and $C_M = 1$. Each epoch is assigned to one of five sleep stages, i.e., $\mathcal{Y} = \{W, N1, N2, N3, REM\}$.

Since sleep stages exhibit clear temporal continuity and transition patterns, we adopt a many-to-one contextual input format. Given a context length $L = 2r + 1$, the input sequence centered at the i-th epoch is constructed as $\mathbf{X}_i = \{\mathbf{x}_{i-r}, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_{i+r}\}$. The model predicts only the sleep stage of the center epoch, formulated as $\hat{\mathbf{p}}_i = f_\Theta(\mathbf{X}_i)$ and $\hat{y}_i = \arg\max_{c \in \mathcal{Y}} \hat{\mathbf{p}}_{i,c}$, where f_Θ denotes EC-SleepMamba and $\hat{\mathbf{p}}_i \in \mathbb{R}^5$ is the predicted probability distribution over the five classes.

Figure 1 illustrates the overall architecture of EC-SleepMamba. Dual-Scale CNNs first extract modality-specific local tokens. Local Bi-Mamba and Cross-Mamba then model intra-epoch dependencies, while the EEG-guided Interaction Gate conditionally modulates EOG/EMG branches using EEG-derived context. Finally, Global Bi-Mamba captures inter-epoch stage transitions and predicts the center epoch.

3.1 Dual-Scale CNN

Clinical sleep staging relies on both long rhythmic patterns (e.g., slow waves, spindles) and transient waveforms (e.g., K-complexes, eye movements). To preserve such multi-scale information without excessive temporal downsampling, we design a modality-specific Dual-Scale CNN (see Figure 2 and Table 1). Inspired by the dual-stream convolutional architecture of SleepPrintNet (Jia et al., 2020), this module mitigates excessive downsampling to produce longer, finer-grained token sequences for

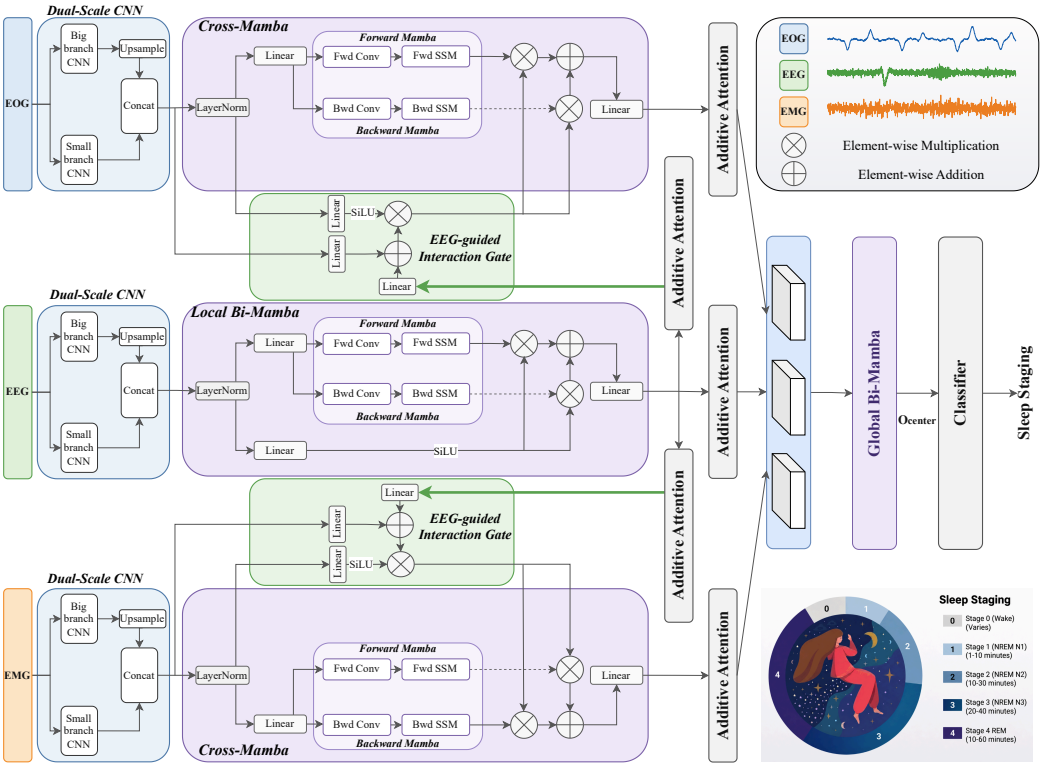

Figure 1: **Overview of our EC-SleepMamba framework.** Modality-specific Dual-Scale CNNs first extract multi-scale physiological features from EEG, EOG, and EMG signals. A Local Bi-Mamba then captures bidirectional intra-epoch EEG dynamics. Guided by EEG-derived sleep-stage semantics, the EEG-guided Interaction Gates modulate the internal selective Z-gates of the EOG and EMG Cross-Mamba branches, enabling directional primary-to-auxiliary interaction during representation encoding. Finally, the fused epoch-level representations are processed by a Global Bi-Mamba to model inter-epoch sleep-stage transitions for final classification.

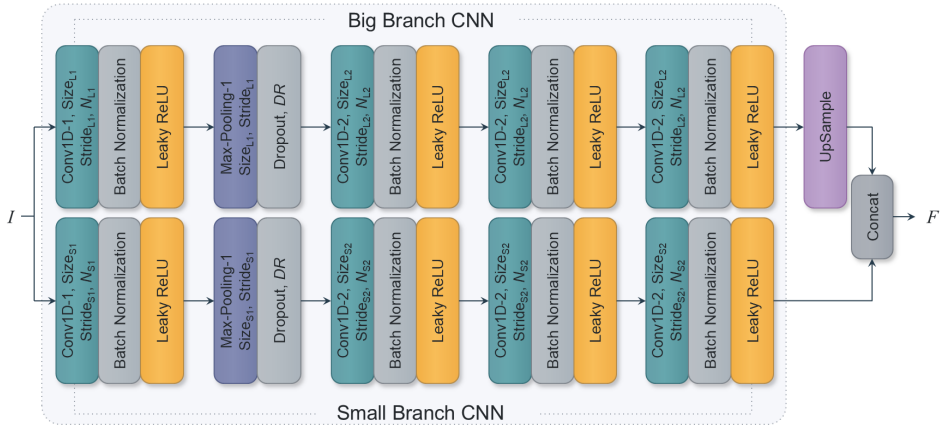

Figure 2: Dual-Scale CNN. Big- and Small-kernel branches are temporally aligned and concatenated; Table 1 lists their convolution and pooling configurations.

the subsequent Mamba scanning. For modality $m \in \{E, O, M\}$, the raw signal $\mathbf{x}_i^m$ is fed into parallel large-kernel (CNN_B^m) and small-kernel (CNN_S^m) branches. The coarse feature map from the large branch is upsampled via transposed convolution to match the temporal resolution of the

Table 1: Convolutional configuration of the proposed Dual-Scale CNN.

Module	Layer	Big branch CNN	Small branch CNN
EEG Dual-Scale CNN	Conv1D-1	$Size_{L1} = 400$, $Stride_{L1} = 50$, $N_{L1} = 64$	$Size_{S1} = 50$, $Stride_{S1} = 6$, $N_{S1} = 64$
	Max-Pooling-1	$Size_{L1} = 4$, $Stride_{L1} = 4$	$Size_{S1} = 8$, $Stride_{S1} = 8$
	Dropout	$DR = 0.5$	$DR = 0.5$
	Conv1D-2	$Size_{L2} = 6$, $Stride_{L2} = 1$, $N_{L2} = 128$	$Size_{S2} = 8$, $Stride_{S2} = 1$, $N_{S2} = 128$
	Upsample	ConvTranspose1d	–
EOG/EMG Dual-Scale CNN	Conv1D-1	$Size_{L1} = 400$, $Stride_{L1} = 50$, $N_{L1} = 32$	$Size_{S1} = 50$, $Stride_{S1} = 6$, $N_{S1} = 32$
	Max-Pooling-1	$Size_{L1} = 4$, $Stride_{L1} = 4$	$Size_{S1} = 8$, $Stride_{S1} = 8$
	Dropout	$DR = 0.5$	$DR = 0.5$
	Conv1D-2	$Size_{L2} = 6$, $Stride_{L2} = 1$, $N_{L2} = 64$	$Size_{S2} = 8$, $Stride_{S2} = 1$, $N_{S2} = 64$
	Upsample	ConvTranspose1d	–

fine branch, and the two are concatenated along the feature dimension:

$$\mathbf{H}_i^{m,0} = \mathrm{CNN}_S^m(\mathbf{x}_i^m) \parallel \mathrm{ConvTranspose1d}\big(\mathrm{CNN}_B^m(\mathbf{x}_i^m)\big), \tag{1}$$

where $\parallel$ denotes channel-wise concatenation. To retain modality-specific physiological signatures before cross-modal interaction, the convolutional weights are not shared across modalities. Let N be the number of tokens per epoch and D be the dimension of the token; thus $\mathbf{H}_i^{m,0} \in \mathbb{R}^{N\times D}$.

3.2 Intra-Epoch Local Bi-Mamba

Although the Dual-Scale CNN captures local waveforms, it cannot model long-range dependencies among sparsely distributed events occurring far apart within an epoch, such as spindles and K-complexes. We therefore adopt the Bidirectional Mamba (Bi-Mamba) block as the intra-epoch backbone, exploiting its linear-complexity selective scan (Gu & Dao, 2024). Given a token sequence $\mathbf{X} \in \mathbb{R}^{N\times D}$, a Bi-Mamba block first normalizes the input and projects it into the state-space branch $\mathbf{U}$ and the internal Z-gate $\mathbf{Z}$:

$$\overline{\mathbf{X}} = \mathrm{LN}(\mathbf{X}), \quad \mathbf{U} = \overline{\mathbf{X}}\mathbf{W}_u, \quad \mathbf{Z} = \overline{\mathbf{X}}\mathbf{W}_z. \tag{2}$$

To capture intra-epoch bidirectional context that allows each token to relate to physiological events (e.g., spindles and K-complexes) occurring at any temporal position within the epoch, we construct forward and backward state-branch inputs:

$$\mathbf{U}_f = \mathbf{U}, \qquad \mathbf{U}_b = \mathrm{Rev}(\mathbf{U}), \tag{3}$$

with $\mathrm{Rev}(\cdot)$ reversing the sequence dimension. Each direction is processed by an independent 1D convolution and SSM:

$$\begin{aligned} \mathbf{Y}_f &= \mathrm{SSM}_f\big(\mathrm{SiLU}(\mathrm{Conv1D}_f(\mathbf{U}_f))\big), \\ \mathbf{Y}_b &= \mathrm{Rev}\big(\mathrm{SSM}_b\big(\mathrm{SiLU}(\mathrm{Conv1D}_b(\mathbf{U}_b))\big)\big). \end{aligned} \tag{4}$$

The forward and backward outputs are summed and gated by the same Z-gate:

$$\mathbf{G} = \mathrm{SiLU}(\mathbf{Z}), \qquad \mathrm{BiMamba}(\mathbf{X}) = \big[(\mathbf{Y}_f + \mathbf{Y}_b) \odot \mathbf{G}\big]\mathbf{W}_o, \tag{5}$$

where $\odot$ is element-wise multiplication and $\mathbf{W}_o$ the output projection. In EC-SleepMamba, the EEG branch applies a Local Bi-Mamba directly to the Dual-Scale CNN tokens:

$$\mathbf{H}_i^E = \mathrm{BiMamba}_{\mathrm{local}}\big(\mathbf{H}_i^{E,0}\big), \tag{6}$$

yielding a context-rich EEG representation within epoch i.

3.3 EEG-Guided Interaction Gate

Different sleep stages rely on different physiological cues. N2 and N3 are primarily defined by EEG events. EOG provides eye-movement evidence for W, N1, and REM, whereas chin EMG provides muscle-tone evidence, especially for REM (Berry et al., 2012; Keenan & Hirshkowitz, 2011). To operationalize this asymmetric clinical prior, we use EEG-derived brain-state context to modulate information flow inside the auxiliary Cross-Mamba branches.

EEG context vector First, we summarize the EEG token sequence into a global context vector using additive attention (Bahdanau et al., 2015):

$$\begin{aligned} \mathbf{u}_{i,n}^{E} &= \tanh(\mathbf{W}_c \mathbf{h}_{i,n}^{E} + \mathbf{b}_c), \qquad \alpha_{i,n}^{E} = \frac{\exp(\mathbf{v}_c^{\top} \mathbf{u}_{i,n}^{E})}{\sum_{j=1}^{N} \exp(\mathbf{v}_c^{\top} \mathbf{u}_{i,j}^{E})}, \\ \mathbf{c}_i^{E} &= \sum_{n=1}^{N} \alpha_{i,n}^{E}\, \mathbf{h}_{i,n}^{E}, \end{aligned} \tag{7}$$

where $\mathbf{h}_{i,n}^{E}$ is the n-th token of $\mathbf{H}_i^{E}$, and $\mathbf{W}_c, \mathbf{b}_c, \mathbf{v}_c$ are learnable parameters. The proposed gate uses epoch-level EEG context rather than tokenwise cross-modal alignment. We therefore aggregate the EEG tokens into the compact vector $\mathbf{c}_i^{E}$. This vector provides a shared condition for the auxiliary gates and avoids the overhead of pairwise token alignment.

Gate generation For an auxiliary modality $a \in \{O, M\}$, the gate generator combines each pre-normalization auxiliary token with the EEG context:

$$\mathbf{g}_{i,n}^{a} = \tanh\left(\mathbf{W}_{\text{aux}}^{a} \mathbf{h}_{i,n}^{a,0} + \mathbf{W}_{\text{ctx}}^{a} \mathbf{c}_i^{E} + \mathbf{b}_r^{a}\right), \tag{8}$$

where $\mathbf{h}_{i,n}^{a,0}$ is the n-th auxiliary token before the LayerNorm used by the corresponding Cross-Mamba encoder. In contrast, the state and Z branches of Cross-Mamba are computed from $\overline{\mathbf{H}}_i^{a,0} = \mathrm{LN}(\mathbf{H}_i^{a,0})$. The complete EEG-guided gate sequence is $\mathbf{G}_i^{a} = [\mathbf{g}_{i,1}^{a}, \ldots, \mathbf{g}_{i,N_a}^{a}]$.

Integration into Cross-Mamba Internally, Cross-Mamba follows the same Bi-Mamba computations as defined in Equations (2)–(5) to obtain its own Z-gate $\mathbf{Z}_i^{a}$ and bidirectional SSM outputs $\mathbf{Y}_{i,f}^{a}, \mathbf{Y}_{i,b}^{a}$. The EEG-guided gate then multiplicatively interacts with the internal Z-gate to form an effective EEG-conditioned selective gate:

$$\widetilde{\mathbf{G}}_i^{a} = \mathrm{SiLU}(\mathbf{Z}_i^{a}) \odot \mathbf{G}_i^{a}. \tag{9}$$

The final output of the auxiliary Cross-Mamba is:

$$\mathbf{H}_i^{a} = \left[(\mathbf{Y}_{i,f}^{a} + \mathbf{Y}_{i,b}^{a}) \odot \widetilde{\mathbf{G}}_i^{a}\right] \mathbf{W}_o^{a}. \tag{10}$$

This design enables context-dependent modulation of auxiliary information. It can attenuate fluctuations that are inconsistent with the EEG-derived context while preserving diagnostically useful eye-movement and muscle-tone cues. The gate is learned end-to-end under the sleep-staging objective. Ground-truth stage labels are not provided as inputs to the gate, and neither auxiliary modality is subject to hard-coded suppression.

3.4 Global Bi-Mamba and Sleep Staging

After intra-epoch modeling, we compress each token sequence into an epoch-level representation using modality-specific additive attention:

$$\begin{aligned} \mathbf{u}_{i,n}^{m} &= \tanh(\mathbf{W}_a^{m} \mathbf{h}_{i,n}^{m} + \mathbf{b}_a^{m}), \qquad \gamma_{i,n}^{m} = \frac{\exp((\mathbf{q}_a^{m})^{\top} \mathbf{u}_{i,n}^{m})}{\sum_{j=1}^{N} \exp((\mathbf{q}_a^{m})^{\top} \mathbf{u}_{i,j}^{m})}, \\ \mathbf{e}_i^{m} &= \sum_{n=1}^{N} \gamma_{i,n}^{m}\, \mathbf{h}_{i,n}^{m}, \quad m \in \{E, O, M\}. \end{aligned} \tag{11}$$

The three vectors are concatenated: $\mathbf{e}_i = [\mathbf{e}_i^{E} \| \mathbf{e}_i^{O} \| \mathbf{e}_i^{M}]$. To capture stage-transition patterns, the epoch-level sequence within the context window $\mathbf{E}_i = [\mathbf{e}_{i-r}, \ldots, \mathbf{e}_i, \ldots, \mathbf{e}_{i+r}] \in \mathbb{R}^{L \times d_e}$ is processed by Global Bi-Mamba. The Global Bi-Mamba employs the identical Bi-Mamba computation defined in Equations (2)–(5); the only difference from the Local Bi-Mamba is that it operates on epoch-level representations instead of intra-epoch tokens:

$$\mathbf{O}_i^{G} = \mathrm{BiMamba}_{\text{global}}(\mathbf{E}_i). \tag{12}$$

Finally, the center position $\mathbf{o}_{\text{center}} = \mathbf{O}_i^{G}[r{+}1]$ is classified:

$$\hat{\mathbf{p}}_i = \mathrm{Softmax}\big(\mathbf{W}_2\, \delta(\mathbf{W}_1 \mathbf{o}_{\text{center}} + \mathbf{b}_1) + \mathbf{b}_2\big), \tag{13}$$

where $\delta(\cdot)$ is the ReLU activation. This many-to-one design lets the prediction benefit from both the epoch's own multimodal representation and the sequential context.

Table 2: Performance on Sleep-EDF-39 and Sleep-EDF-153. Best values are bold; the SleepTransformer-153 result excludes SHHS pretraining.

Method	Sleep-EDF-39 dataset							Sleep-EDF-153 dataset						
	Overall results		F1-score for each class					Overall results		F1-score for each class				
	MF1	ACC	Wake	N1	N2	N3	REM	MF1	ACC	Wake	N1	N2	N3	REM
SleepTransformer (Phan et al., 2022b)	67.4	77.7	80.5	19.7	84.8	81.2	70.9	74.3	81.4	91.7	40.4	84.3	77.9	77.2
TinySleepNet (Supratak & Guo, 2020)	76.8	83.1	90.2	43.2	86.2	81.7	82.8	74.7	80.5	90.4	47.2	83.7	74.4	77.5
DeepSleepNet (Supratak et al., 2017)	78.0	83.2	88.6	46.9	86.9	**85.9**	81.9	76.6	82.2	91.1	45.4	**86.3**	**79.8**	80.3
SalientSleepNet (Jia et al., 2021)	78.9	83.9	86.7	51.0	87.3	83.9	85.5	77.3	82.8	92.9	49.9	84.5	75.4	83.8
SeqSleepNet (Phan et al., 2019)	79.3	85.2	90.6	47.4	87.6	**85.9**	85.1	76.9	82.7	91.6	47.6	85.0	76.8	83.3
EC-SleepMamba	**82.3**	**86.5**	**91.1**	**57.7**	**88.7**	84.9	**89.0**	**79.7**	**84.0**	**93.3**	**55.5**	85.1	78.3	**86.3**

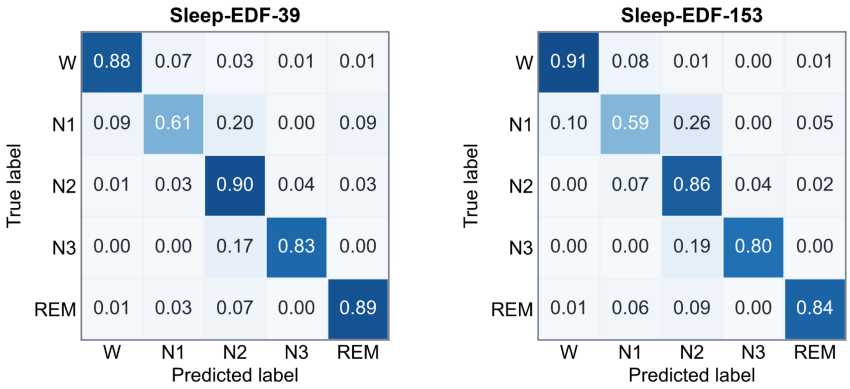

Figure 3: Normalized confusion matrices of EC-SleepMamba on Sleep-EDF-39 (left) and Sleep-EDF-153 (right).

4 Experimental Results and Analysis

The proposed model is implemented in PyTorch, and subject-level 10-fold cross-validation is adopted. The model is trained end-to-end using the Adam optimizer (Kingma & Ba, 2015) with a learning rate of 0.001, Adam parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, a batch size of 128, and 50 training epochs. At a sampling rate of 100 Hz, the input dimension of each epoch is 3000. The context window length is set to $L = 13$, i.e., the six epochs before and after the center epoch are used to predict the sleep stage of the center epoch. Considering that classes such as N1 contain fewer samples, we use weighted cross-entropy, with loss weights for the five classes $\{W, N1, N2, N3, REM\}$ set to $\{1.0, 1.30, 1.10, 1.0, 1.0\}$. Evaluation metrics include Accuracy (ACC), Macro-F1 (MF1), and per-class F1-score.

4.1 Datasets and Preprocessing

We evaluate model performance on Sleep-EDF-39 and Sleep-EDF-153, both derived from the Sleep-EDF Expanded database on PhysioNet (Kemp et al., 2000; Goldberger et al., 2000). After the preprocessing described below, Sleep-EDF-39 contains 42,308 sleep epochs from 20 healthy subjects, while Sleep-EDF-153 contains 195,479 sleep epochs from 78 healthy subjects. Continuous PSG signals are segmented into non-overlapping 30-s epochs. After discarding movement and unknown/unscored epochs, the remaining Rechtschaffen–Kales sleep labels were $\{W, N1, N2, N3, N4, REM\}$ (Rechtschaffen & Kales, 1968). Following the AASM sleep staging standard (Iber et al., 2007), N3 and N4 are merged into a single N3 stage, resulting in a five-class task $\{W, N1, N2, N3, REM\}$. Following the recommendation of Imtiaz & Rodriguez-Villegas (2014),

redundant wake epochs before and after the in-bed period are removed, and only in-bed recordings are retained for model training and testing. The model input channels include EEG Fpz-Cz, EEG Pz-Oz, EOG horizontal, and chin EMG. After preprocessing, every input channel is represented at 100 Hz, so each 30-s epoch contains 3000 samples per channel. Finally, to reduce physiological amplitude differences across subjects and channels, z-score normalization is independently applied to each channel of each subject.

4.2 Comparison with Competing Methods

We compare EC-SleepMamba with five representative automatic sleep-staging baselines, including DeepSleepNet (Supratak et al., 2017), SeqSleepNet (Phan et al., 2019), SleepTransformer (Phan et al., 2022b), TinySleepNet (Supratak & Guo, 2020), and SalientSleepNet (Jia et al., 2021). DeepSleepNet combines CNN feature extraction with Bi-LSTM stage-transition modeling. SeqSleepNet formulates sleep staging as a sequence-to-sequence task. SleepTransformer introduces self-attention for contextual modeling. TinySleepNet is a lightweight CNN-RNN model, while SalientSleepNet improves staging performance through multimodal salient-wave detection and multi-scale transition modeling. For a controlled comparison, all baselines are reproduced using their publicly available implementations under the same subject-wise data splits and evaluation protocol as EC-SleepMamba. Method-specific architectural settings follow the corresponding public implementations unless otherwise stated. BiT-MamSleep (Zhou et al., 2024) is closely related to our work because it also employs bidirectional Mamba for automatic sleep staging. However, since a public implementation was not available for controlled reproduction, we discuss it in the related work but do not include it in the unified rerun comparison in Table 2.

As shown in Table 2, EC-SleepMamba achieves the highest ACC and MF1 on both datasets. Relative to the strongest baselines, MF1 and ACC improve by 3.0 and 1.3 percentage points on Sleep-EDF-39. The corresponding gains on Sleep-EDF-153 are 2.4 and 1.2 points. The class-wise improvements are particularly clear for N1 and REM. SalientSleepNet already performs stage-adaptive multimodal emphasis (Jia et al., 2021). The further gains of EC-SleepMamba, together with the controlled gate ablations below, support moving multimodal adaptation from post-encoding feature emphasis to EEG-conditioned auxiliary representation formation. Figure 3 presents the normalized confusion matrices on Sleep-EDF-39 and Sleep-EDF-153. W, N2, N3, and REM retain high diagonal rates on both datasets, whereas N1 remains the most difficult stage (0.61/0.59). Its errors are concentrated in the adjacent W and N2 stages, with virtually no confusion with N3. A second directional pattern appears between N2 and N3: N3 is more often classified as N2 (0.17/0.19) than N2 as N3 (0.04/0.04), consistent with the threshold-based boundary between these two EEG-dominant stages. These patterns should be interpreted separately. The N2–N3 asymmetry mainly reflects ambiguity in EEG slow-wave evidence, whereas the N1 and REM results are more directly relevant to auxiliary EOG/EMG cues. Together with the gate ablations in Table 3, the improved N1 and REM F1 scores are consistent with the intended role of EEG-conditioned auxiliary regulation. Overall, the remaining errors are concentrated between physiologically adjacent stages rather than being broadly distributed across unrelated classes.

4.3 Ablation Study

We conduct main-module and fine-grained gate ablations to assess the Dual-Scale CNN, intra-epoch Mamba branches, EEG-guided Interaction Gate, and Global Bi-Mamba. **Main Module Ablation** The main-module ablation includes five variants: (1) **w/o Dual-Scale CNN**: the dual-scale front-end is replaced with a single-scale CNN; (2) **w/o Local Bi-Mamba and Cross-Mamba**: intra-epoch Mamba blocks in all modality branches are removed, and the EEG-guided gate is consequently disabled; (3) **w/o Auxiliary Cross-Mamba**: the EEG Local Bi-Mamba is retained, but the EOG/EMG Cross-Mamba branches are removed; (4) **w/o EEG-guided Interaction Gate**: the auxiliary Cross-Mamba blocks are retained, but the EEG-guided gates inside them are removed; and (5) **w/o Global Bi-Mamba**: inter-epoch Global Bi-Mamba is removed.

The complete model achieves the best overall MF1 and ACC, and every removal lowers both metrics. Replacing the Dual-Scale CNN reduces MF1 to 79.4%, while removing all intra-epoch Mamba branches yields 80.7%. Retaining the EEG Local Bi-Mamba but removing the auxiliary Cross-Mamba branches gives 81.6%, showing that EEG remains informative while regulated EOG/EMG

Table 3: Main ablation results. Bold values indicate the best result for each metric.

Method	Overall		Per-class F1				
	MF1	ACC	W	N1	N2	N3	REM
w/o Dual-Scale CNN	79.4	83.5	86.4	54.4	86.7	85.0	84.3
w/o Local Bi-Mamba and Cross-Mamba	80.7	84.6	86.9	57.0	87.3	85.9	86.2
w/o Auxiliary Cross-Mamba	81.6	85.7	89.6	56.3	88.0	**86.5**	87.5
w/o EEG-Guided Interaction Gate	80.4	84.5	87.5	55.1	87.4	85.6	86.5
w/o Global Bi-Mamba	78.6	83.3	88.7	51.5	86.7	85.1	80.8
Complete (Ours)	**82.3**	**86.5**	**91.1**	**57.7**	**88.7**	84.9	**89.0**

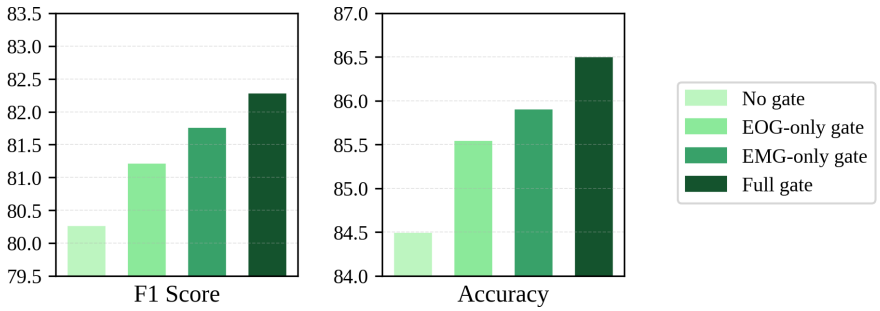

Figure 4: Fine-grained gate ablation: MF1 (left) and ACC (right).

representations provide additional value. Removing only the EEG-guided gate drops MF1 to 80.4%, indicating that retaining auxiliary Cross-Mamba without EEG conditioning does not by itself guarantee effective fusion. The largest decrease occurs without Global Bi-Mamba, supporting the importance of inter-epoch context.

Fine-Grained Gate Ablation We further compare no gate, single-modality EOG- or EMG-only gating, and full gating of both auxiliary branches.

As Figure 4 shows, either single-modality gate improves ACC and MF1 over no gate, indicating that EEG conditioning benefits both auxiliary signals. Full gating performs best, supporting the conclusion that EEG-regulated EOG and EMG provide complementary evidence at the stage. The improvement over the no-gate variant further indicates that the performance gain stems from EEG-conditioned regulation rather than auxiliary encoding alone, supporting the proposed directional primary-to-auxiliary interaction.

5 Conclusion

In this work, we present EC-SleepMamba, an EEG-guided cross-modal state-space framework for robust multimodal sleep staging. Unlike conventional symmetric fusion methods, EC-SleepMamba explicitly models the asymmetric clinical roles of EEG, EOG, and EMG by using EEG-derived sleep-stage semantics to regulate the internal selective gates of the auxiliary Cross-Mamba branches. Moreover, the hierarchical combination of Local Bi-Mamba, Cross-Mamba, and Global Bi-Mamba jointly captures bidirectional intra-epoch physiological dynamics and inter-epoch sleep-stage transitions with linear computational complexity in sequence length. Experiments on Sleep-EDF-39 and Sleep-EDF-153 show that EC-SleepMamba consistently outperforms competing approaches and improves the recognition of highly confusable sleep stages, demonstrating its effectiveness for sleep staging.

AI USE STATEMENT

In this work, we used generative AI tools to assist with manuscript writing and language polishing, as well as research and implementation support. We have reviewed all AI-assisted work and revised or verified the generated content where necessary. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We have made efforts to ensure the reproducibility of this work. The architecture and key components of EC-SleepMamba are described in the methodology section, while the datasets, preprocessing procedures, cross-validation protocol, evaluation metrics, and training configurations are detailed in the experimental section. The source code will be released upon acceptance.

REFERENCES

Fernando Andreotti, Huy Phan, Navin Cooray, Christine Lo, Michele T. M. Hu, and Maarten De Vos. Multichannel sleep stage classification and transfer learning using convolutional neural networks. In *Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 171–174. IEEE, 2018. doi: 10.1109/EMBC.2018.8512214.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations*, 2015. URL https://mlanthology.org/iclr/2015/bahdanau2015iclr-neural/.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, 2018.

Richard B. Berry, Rita Brooks, Charlene E. Gamaldo, Susan M. Harding, Carole L. Marcus, and Bradley V. Vaughn. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0*. American Academy of Sleep Medicine, Darien, IL, USA, 2012.

Aydar Bulatov, Yuri Kuratov, and Mikhail Burtsev. Recurrent memory transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 11079–11091, 2022.

Stanislas Chambon, Mathieu N. Galtier, Pierrick J. Arnal, Gilles Wainrib, and Alexandre Gramfort. 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, 2018. doi: 10.1109/TNSRE.2018.2813138.

Luigi Fiorillo, Alessandro Puiatti, Michela Papandrea, Pietro-Luca Ratti, Paolo Favaro, Corinne Roth, Panagiotis Bargiotas, Claudio L. Bassetti, and Francesca D. Faraci. Automated sleep scoring: A review of the latest approaches. *Sleep Medicine Reviews*, 48:101204, 2019. doi: 10.1016/j.smrv.2019.07.007.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. doi: 10.1161/01.CIR.101.23.e215.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling (COLM)*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *The Tenth International Conference on Learning Representations (ICLR)*, 2022.

Conrad Iber, Sonia Ancoli-Israel, Andrew L. Chesson, and Stuart F. Quan. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine, Westchester, IL, USA, 2007.

Syed Anas Imtiaz and Esther Rodriguez-Villegas. Recommendations for performance assessment of automatic sleep staging algorithms. In *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2014)*, pp. 5044–5047. IEEE, 2014. doi: 10.1109/EMBC.2014.6944758.

Institute of Medicine. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press, Washington, DC, USA, 2006. doi: 10.17226/11617.

Ziyu Jia, Xiyang Cai, Gaoxing Zheng, Jing Wang, and Youfang Lin. SleepPrintNet: A multivariate multimodal neural network based on physiological time-series for automatic sleep staging. *IEEE Transactions on Artificial Intelligence*, 1(3):248–257, 2020. doi: 10.1109/TAI.2021.3060350.

Ziyu Jia, Youfang Lin, Jing Wang, Xuehui Wang, Peiyi Xie, and Yingbin Zhang. SalientSleepNet: Multimodal salient wave detection network for sleep staging. In Zhi-Hua Zhou (ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 2614–2620. International Joint Conferences on Artificial Intelligence Organization, 8 2021. doi: 10.24963/ijcai.2021/360. URL https://doi.org/10.24963/ijcai.2021/360.

Sharon Keenan and Max Hirshkowitz. Monitoring and staging human sleep. In Meir H. Kryger, Thomas Roth, and William C. Dement (eds.), *Principles and Practice of Sleep Medicine*, pp. 1602–1609. Elsevier Saunders, St. Louis, MO, USA, 5th edition, 2011. doi: 10.1016/B978-1-4160-6645-3.00141-9.

Bob Kemp, Aeilko H. Zwinderman, Bert Tuk, Hilbert A. C. Kamphuisen, and Josefien J. L. Oberye. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering*, 47(9):1185–1194, 2000. doi: 10.1109/10.867928.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the Third International Conference on Learning Representations (ICLR 2015)*, 2015.

Maurice M. Ohayon. Epidemiological overview of sleep disorders in the general population. *Sleep Medicine Research*, 2(1):1–9, 2011. doi: 10.17241/smr.2011.2.1.1.

Mathias Perslev, Michael Hejselbak Jensen, Sune Darkner, Poul Jørgen Jennum, and Christian Igel. U-Time: A fully convolutional network for time series segmentation applied to sleep staging. In *Advances in Neural Information Processing Systems*, volume 32, pp. 4415–4426, 2019.

Huy Phan, Fernando Andreotti, Navin Cooray, Oliver Y. Chén, and Maarten De Vos. SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(3):400–410, 2019. doi: 10.1109/TNSRE.2019.2896659.

Huy Phan, Oliver Y. Chén, Minh C. Tran, Philipp Koch, Alfred Mertins, and Maarten De Vos. XSleepNet: Multi-view sequential model for automatic sleep staging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5903–5915, 2022a. doi: 10.1109/TPAMI.2021.3070057.

Huy Phan, Kaare Mikkelsen, Oliver Y. Chén, Philipp Koch, Alfred Mertins, and Maarten De Vos. SleepTransformer: Automatic sleep staging with interpretability and uncertainty quantification. *IEEE Transactions on Biomedical Engineering*, 69(8):2456–2467, 2022b. doi: 10.1109/TBME.2022.3147187.

Allan Rechtschaffen and Anthony Kales (eds.). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. UCLA Brain Information Service, Brain Research Institute, Los Angeles, CA, USA, 1968.

Marina Ronzhina, Oto Janoušek, Jana Kolářová, Marie Nováková, Petr Honzík, and Ivo Provazník. Sleep scoring using artificial neural networks. *Sleep Medicine Reviews*, 16(3):251–263, 2012. doi: 10.1016/j.smrv.2011.06.003.

Richard S. Rosenberg and Susan Van Hout. The american academy of sleep medicine inter-scorer reliability program: Sleep stage scoring. *Journal of Clinical Sleep Medicine*, 9(1):81–87, 2013. doi: 10.5664/jcsm.2350.

Qi Shen, Junchang Xin, Bing Tian Dai, Shudi Zhang, and Zhiqiong Wang. Robust sleep staging over incomplete multimodal physiological signals via contrastive imagination. In *Advances in Neural Information Processing Systems*, volume 37, pp. 112025–112049. Curran Associates, Inc., 2024. doi: 10.52202/079017-3557.

Akara Supratak and Yike Guo. TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel EEG. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 641–644. IEEE, 2020. doi: 10.1109/EMBC44109.2020.9176741.

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. doi: 10.1109/TNSRE.2017.2721116.

Orestis Tsinalis, Paul M. Matthews, Yike Guo, and Stefanos Zafeiriou. Automatic sleep stage scoring with single-channel EEG using convolutional neural networks, 2016.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017.

Jing Wang, Xuehui Wang, Xiaojun Ning, Youfang Lin, Huy Phan, and Ziyu Jia. Subject-adaptation salient wave detection network for multimodal sleep stage classification. *IEEE Journal of Biomedical and Health Informatics*, 29(3):2172–2184, 2025. doi: 10.1109/JBHI.2024.3512584.

Weilian Zhou, Sei-Ichiro Kamata, Haipeng Wang, and Xi Xue. Multiscanning-based RNN–transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:5512319, 2023. doi: 10.1109/TGRS.2023.3277014.

Xinliang Zhou, Yuzhe Han, Zhisheng Chen, Chenyu Liu, Yi Ding, Ziyu Jia, and Yang Liu. BiT-MamSleep: Bidirectional temporal mamba for EEG sleep staging. *arXiv preprint arXiv:2411.01589*, 2024. doi: 10.48550/arXiv.2411.01589. URL `https://arxiv.org/abs/2411.01589`.