CalibSleep: Cross-Modal Calibration and Rule-Aware Learning for Automatic Sleep Stage Classification

Xiang Chen[1], Jia Shu[1], and Zheyuan Wang[1]*

Shanghai Jiao Tong University, Shanghai, China
zheyuan_wang@sjtu.edu.cn

Abstract. Automatic sleep staging is a fundamental task in sleep health assessment and clinical diagnosis, yet it remains challenging due to insufficient exploitation of multimodal complementarity, heterogeneous physiological signals, and complex sleep stage transition patterns. This paper proposes **CalibSleep**, a dual-stream sleep stage classification framework that jointly models time-domain EEG/EOG signals and their corresponding time–frequency representations. CalibSleep employs modality-specific encoders to capture complementary temporal and spectral features, and introduces a cross-modal calibration module to explicitly align heterogeneous representations and adaptively balance modality contributions. Furthermore, physiological prior knowledge of sleep stage transitions is incorporated through a rule-aware classification strategy, improving prediction stability and clinical consistency without relying on complex long-range temporal modeling. Experiments on an in-house clinical dataset from Shanghai Jiao Tong University Affiliated Sixth People's Hospital and the public SHHS1 dataset demonstrate that CalibSleep achieves accuracies of 89.6% and 88.3%, respectively, outperforming existing methods. Ablation studies further validate the effectiveness of each proposed component. These results indicate that CalibSleep provides an effective and well-generalizable solution for automatic sleep stage classification. The code is publicly available at https://github.com/Cx3300903/CalibSleep.

Keywords: Sleep Stage Classification· Time–Frequency Representation· Cross-Modal Calibration· Rule-Aware Learning.

1 Introduction

Sleep stage classification is a core task for sleep health monitoring and clinical diagnosis, playing an indispensable role in sleep disorder early screening [1], neurodegenerative disease progression tracking [2], and personal health management [3]. In clinical practice, sleep stages are manually annotated from PSG recordings, including EEG, EOG, EMG signals, by segmenting signals into 30-second epochs according to the AASM guidelines [3,4]. However, the large volume

* Corresponding author.

of PSG data makes manual sleep staging time-consuming and highly dependent on expert experience [5]. These limitations have motivated research on automated sleep staging methods [6], while early approaches relied on handcrafted features and traditional machine learning techniques with limited scalability [7].

With the development of DL, automatic feature learning has been applied to sleep stage classification [13,14]. However, existing methods still exhibit two major limitations in multimodal feature mining: **over-reliance on single-channel signals** and **one-sided domain-specific feature extraction**. Several studies perform sleep staging using only single-channel EEG [8,9] or ECG signals [10], which capture only local physiological patterns and fail to model global sleep dynamics. Even when multi-channel data are used, most approaches focus exclusively on either raw time-domain signals or time–frequency representations. For example, Wang *et al.* [11] treated signal waveforms as images with limited performance on N1 and N3 stages, while FlexibleSleepNet [12] achieved 86.9% accuracy using fused time–frequency features of EEG and EOG. Since time-domain signals and time–frequency images describe complementary temporal dynamics and spectral energy distributions, neglecting their joint modeling inevitably limits the completeness and discriminative capacity of learned representations.

Beyond feature extraction, the **feature alignment** and **fusion mechanisms** between time-domain signals and time–frequency images remain insufficiently explored. Some studies [15] simply concatenated or superimposed features from different domains, resulting in limited performance gains. XSleepNet [16] learned joint representations while preserving modality-specific features, and SEN-DAL [17] adopted domain adversarial learning to address multimodal heterogeneity. Despite these advances, existing methods still lack explicit modeling of **heterogeneous feature conflicts** and inter-modal interactions.

In addition to intra-epoch feature modeling, sleep stage classification also relies on capturing the **transition relationships** between adjacent sleep stages. In normal sleep cycles, stages typically evolve sequentially from W to N1, N2, and N3, followed by transitions to REM [4]. Direct transitions across distant stages are observed in clinical practice and are often associated with disorders or annotation noise. However, most existing methods neglect the rationality of stage transitions [18], leading to predictions that violate sleep patterns.

To address these challenges, we propose **CalibSleep**, a calibrated dual-stream framework for automatic sleep stage classification. CalibSleep jointly models raw time-domain EEG/EOG signals and their corresponding time–frequency representations, explicitly calibrates heterogeneous cross-modal features, and incorporates physiological sleep-stage transition priors as a lightweight regularization strategy. The contributions of this work are summarized as follows:

1. We formulate sleep stage classification as a calibrated dual-stream learning problem, where time-domain EEG/EOG waveforms and time–frequency representations are jointly modeled to capture complementary temporal dynamics and spectral sleep signatures.
2. We propose a Cross-Modal Calibration and Fusion (CMC) module that goes beyond naive feature concatenation by explicitly aligning heterogeneous rep-

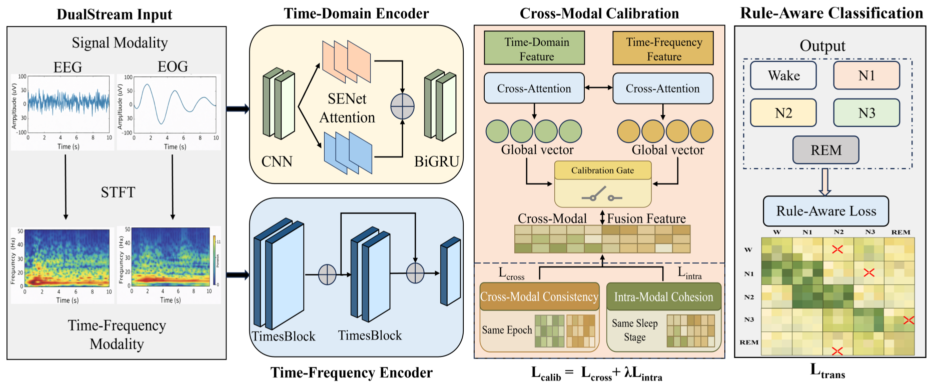

Fig. 1. Architecture of the proposed CalibSleep model.

resentations through bidirectional cross-attention and adaptively balancing modality contributions using a learnable calibration gate.

3. We introduce a rule-aware transition regularization strategy that incorporates physiological sleep-stage transition priors as a soft training constraint, improving prediction consistency and clinical plausibility without relying on computationally expensive long-sequence modeling.

2 Method

2.1 Overall Architecture of CalibSleep

The overall architecture of CalibSleep is illustrated in Fig. 1. The framework processes synchronized multi-channel EEG/EOG signals through two parallel branches: a time-domain branch that extracts local waveform and temporal dynamics from 30-second epochs, and a time–frequency branch that captures rhythmic and periodic structures from corresponding time–frequency representations. The resulting heterogeneous features are integrated by a cross-modal calibration and fusion module, which performs bidirectional cross-attention for semantic alignment and applies adaptive calibration to balance modality contributions. The fused representation is finally fed into a rule-aware classification head to produce sleep stage probabilities under soft stage transition regularization.

2.2 Modality-Specific Feature Encoders

Time-Domain Signal Encoder. The time-domain encoder consists of a multi-scale 1D convolutional network followed by a bidirectional GRU (BiGRU). The convolutional layers capture local waveform patterns at different temporal scales, while a SENet-style [20] channel attention mechanism is employed to adaptively reweight EEG and EOG channels. The channel-weighted features are then fed

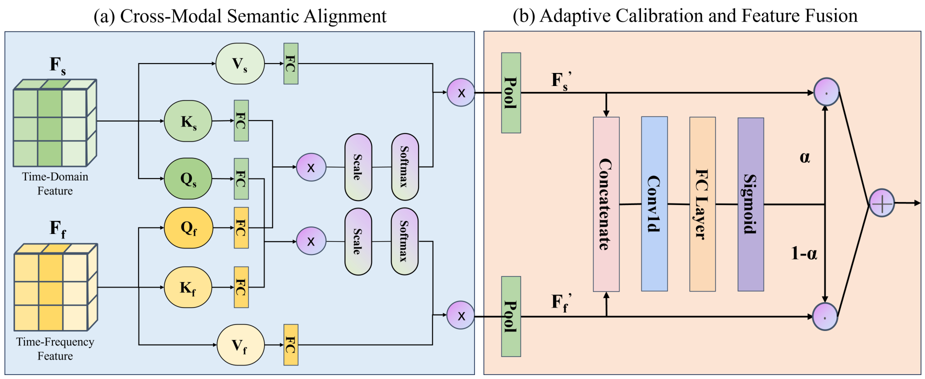

Fig. 2. Architecture of the cross-modal calibration and fusion module.

into the BiGRU [19] to model intra-epoch temporal dependencies and generate an epoch-level representation:

$$\mathbf{H}_s = \mathcal{C}(\mathbf{X}_s), \quad \tilde{\mathbf{H}}_s = \mathbf{H}_s \odot \sigma(\mathbf{W}\,\mathrm{GAP}(\mathbf{H}_s)), \quad \mathbf{f}_s = \mathbf{W}_p\,\mathcal{R}(\tilde{\mathbf{H}}_s)_{\mathrm{last}}, \tag{1}$$

where $\mathcal{C}(\cdot)$ denotes the convolutional encoder, $\mathrm{GAP}(\cdot)$ is temporal global average pooling, $\sigma(\cdot)$ is the Sigmoid function, $\odot$ denotes channel-wise weighting, and $\mathcal{R}(\cdot)$ represents the BiGRU operation.

Time–Frequency Encoder Based on TimesNet. In the time–frequency branch, the short-time Fourier transform (STFT) is applied to obtain a time–frequency representation, which is organized as a multivariate temporal sequence:

$$\mathbf{X}_f = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N], \quad \mathbf{z}_n \in \mathbb{R}^F, \tag{2}$$

where F denotes the frequency dimension and N denotes the number of time steps. A TimesNet [21] encoder is adopted to capture latent rhythmic and periodic structures in the time–frequency domain, producing high-level spectral–temporal features. To facilitate subsequent cross-modal calibration, features from both branches are projected into a shared embedding space:

$$\mathbf{F}_f = \mathcal{T}(\mathbf{X}_f), \quad \tilde{\mathbf{F}}_s = \mathbf{F}_s\mathbf{W}_s, \quad \tilde{\mathbf{F}}_f = \mathbf{F}_f\mathbf{W}_f, \tag{3}$$

where $\mathcal{T}(\cdot)$ denotes the period-aware transformation in TimesNet, and $\mathbf{W}_s$, $\mathbf{W}_f$ are learnable projection matrices. The projected features $\tilde{\mathbf{F}}_s$ and $\tilde{\mathbf{F}}_f$ are subsequently fed into the CMC module for cross-modal fusion.

2.3 Cross-Modal Calibration Fusion

Cross-Modal Semantic Alignment: To mitigate cross-modal semantic discrepancies, the CMC module introduces a bidirectional cross-attention mechanism, which uses time-domain features as queries and time-frequency features as keys and values. The alignment process is expressed as:

$$\mathbf{Q}_s = \tilde{\mathbf{F}}_s \mathbf{W}_Q^s, \mathbf{K}_f = \tilde{\mathbf{F}}_f \mathbf{W}_K^f, \mathbf{V}_f = \tilde{\mathbf{F}}_f \mathbf{W}_V^f, \mathbf{F}'_s = \mathrm{softmax}\left(\frac{\mathbf{Q}_s \mathbf{K}_f^\top}{\sqrt{d}}\right)\mathbf{V}_f. \quad (4)$$

Similarly, by taking time-frequency features as queries and time-domain features as keys and values, we obtain:

$$\mathbf{Q}_f = \tilde{\mathbf{F}}_f \mathbf{W}_Q^f, \mathbf{K}_s = \tilde{\mathbf{F}}_s \mathbf{W}_K^s, \mathbf{V}_s = \tilde{\mathbf{F}}_s \mathbf{W}_V^s, \mathbf{F}'_f = \mathrm{softmax}\left(\frac{\mathbf{Q}_f \mathbf{K}_s^\top}{\sqrt{d}}\right)\mathbf{V}_s. \quad (5)$$

Adaptive Calibration and Feature Fusion: To dynamically balance the contribution of each modality, the CMC module introduces an adaptive calibration mechanism. First, global feature representations are obtained via pooling along the time dimension; subsequently, the calibration coefficients are calculated:

$$\bar{\mathbf{f}}_s = \mathrm{Pool}(\mathbf{F}'_s), \quad \bar{\mathbf{f}}_f = \mathrm{Pool}(\mathbf{F}'_f), \quad \alpha = \sigma\left(\mathbf{w}^\top \left[\bar{\mathbf{f}}_s \parallel \bar{\mathbf{f}}_f\right] + b\right), \quad (6)$$

where $\sigma(\cdot)$ denotes the Sigmoid function. Finally, fused feature expressed as:

$$\mathbf{F}_{\text{fusion}} = \alpha \cdot \mathbf{F}'_s + (1-\alpha) \cdot \mathbf{F}'_f. \quad (7)$$

This mechanism enables the model to adaptively select more reliable modal information, thus avoiding performance degradation caused by fixed-weight fusion.

Contrastive Calibration Loss (Training Phase): To further stabilize the cross-modal alignment process during training, this paper introduces a contrastive calibration loss to impose constraints on feature representations at the distribution level. For cross-modal features of the same epoch, their latent representations should be consistent; meanwhile, intra-modal constraints are applied to cluster sample features of the same sleep stage. Thus, the cross-modal and intra-modal losses are defined as:

$$\mathcal{L}_{\text{cross}} = \sum_t \left\| \bar{\mathbf{f}}_s^{(t)} - \bar{\mathbf{f}}_f^{(t)} \right\|_2^2, \quad \mathcal{L}_{\text{intra}} = \sum_{m \in \{s,f\}} \sum_{(i,j) \in \mathcal{P}} \left\| \bar{\mathbf{f}}_m^{(i)} - \bar{\mathbf{f}}_m^{(j)} \right\|_2^2, \quad (8)$$

where $\mathcal{P}$ denotes the set of sample pairs sharing the same sleep stage label. The contrastive calibration loss is thus obtained by combining the above components. This loss is only used during the training phase and does not increase the computational complexity in the inference phase.

$$\mathcal{L}_{\text{calib}} = \mathcal{L}_{\text{cross}} + \lambda \mathcal{L}_{\text{intra}}, \quad (9)$$

2.4 Rule-Aware Sleep Stage Classification

After cross-modal calibration and fusion, sleep stage prediction is performed based on the fused representation. A lightweight multi-layer perceptron (MLP) is employed to generate classification logits, followed by a softmax function to obtain the sleep stage probability distribution:

$$\mathbf{z}_t = \mathcal{G}\left(\mathbf{F}_{\text{fusion}}^{(t)}\right), \quad \mathbf{p}_t = \text{softmax}(\mathbf{z}_t), \tag{10}$$

where $\mathbf{p}_t \in \mathbb{R}^K$ and $K = 5$ corresponds to the Wake, N1, N2, N3, and REM.

To encourage physiological consistency in stage transitions, a rule-aware transition regularization term is introduced. A stage transition matrix $M \in [0,1]^{K\times K}$ is constructed based on commonly observed AASM-based sleep-stage transition patterns and hypnogram continuity priors, where M_{ij} denotes the physiological plausibility of transitioning from stage i to stage j. It should be noted that the transition matrix is used as a soft regularization prior rather than a hard post-processing constraint. Therefore, the model is still allowed to predict atypical transitions when supported by signal evidence, while high-confidence physiologically implausible transitions are discouraged during training. For two adjacent epochs, the predicted transition distribution and the corresponding regularization loss are defined as:

$$\mathbf{T}_t = \mathbf{p}_{t-1}\mathbf{p}_t^\top, \quad \mathcal{L}_{\text{trans}} = \sum_{i=1}^{K}\sum_{j=1}^{K} T_t(i,j)\,(1 - M_{ij}). \tag{11}$$

The overall training objective is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_{\text{calib}}\mathcal{L}_{\text{calib}} + \lambda_{\text{trans}}\mathcal{L}_{\text{trans}}, \tag{12}$$

where $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss, $\mathcal{L}_{\text{calib}}$ denotes the cross-modal calibration loss, and $\mathcal{L}_{\text{trans}}$ is the stage transition regularization term.

3 Experiments and Results

3.1 Datasets and Implementation

No.6 Hospital Dataset (Affiliated Sixth People's Hospital of Shanghai Jiao Tong University): It includes PSG data of 358 subjects collected from 2024 to 2025, with EEG C3-A2 and EOG LOC-A2 signals acquired at a sampling rate of 100 Hz. Each 30-second signal segment is labeled with sleep stages in accordance with the AASM guidelines.

SHHS1: It contains PSG data of 5793 participants aged over 40, collected from November 1995 to January 1998. The same EEG and EOG signals were selected and divided into 30-second epochs, with all sleep stage labels following the AASM criteria. In the experiment, SHHS1 was mainly used as an external dataset to evaluate the generalization ability of the model.

Implementation Details: All raw signals were resampled to 100 Hz, with low-quality waveforms removed, and segmented into non-overlapping 30-second epochs following the AASM guidelines. Time-domain signals were normalized to zero mean and unit variance. For the time–frequency branch, short-time Fourier transform (STFT) was applied using a 2-second window with 50% overlap. The No.6 Hospital dataset was split into training, validation, and test sets with a ratio of 7:1.5:1.5. The model was trained using the Adam optimizer with an initial learning rate of 1×10^{-5} and batch size of 32. The loss weights were set to $\lambda_{\text{calib}} = 1.0$ and $\lambda_{\text{trans}} = 0.1$. All experiments were implemented in PyTorch and conducted on an NVIDIA RTX 4090 GPU. Since the calibration and transition losses are only used during training, they do not introduce additional computational overhead during inference.

Table 1. Performance Comparison of Different Input Modality Methods.

Method	Time	TF	No.6 Hospital			SHHS1		
			ACC	MF1	K	ACC	MF1	K
XSleepNet [23]	✓		77.4	73.5	0.77	79.9	74.3	0.74
SleepBoost [24]	✓		81.7	75.6	0.76	81.4	80.1	0.81
CoReSleep [25]	✓		80.3	77.4	0.76	79.1	80.6	0.80
SleepNet [26]		✓	84.3	79.4	0.81	83.9	80.3	0.78
SeqSNet [27]		✓	85.1	82.6	0.82	84.7	82.2	0.81
SFormer [28]		✓	81.1	77.9	0.79	80.7	81.1	0.76
MultiSNet [29]	✓	✓	84.7	82.3	0.84	81.5	80.3	0.81
CFSNet [30]	✓	✓	87.3	84.1	0.85	85.9	84.1	0.82
CalibSleep	✓	✓	**89.6**	**87.7**	**0.87**	**88.3**	**85.9**	**0.85**

3.2 Comparative Results and Analysis

Tab. 1 reports the performance of unimodal and time–frequency multimodal sleep staging models on the No.6 Hospital and SHHS1 datasets, evaluated by accuracy (ACC), macro F1-score (MF1), and Cohen's Kappa (K). CalibSleep achieves the best performance on both datasets, reaching accuracies of 89.6% and 88.3%, respectively. Compared with the best unimodal baselines, CalibSleep improves accuracy by 4.5% on the No.6 Hospital dataset and 3.6% on SHHS1. Compared with the strongest multimodal baseline, CalibSleep further improves accuracy by 2.3% and 2.4% on the two datasets, respectively. These results demonstrate that explicit cross-modal calibration can better exploit complementary time-domain and time–frequency information, leading to improved robustness and cross-dataset generalization.

3.3 Ablation Study

To evaluate the contribution of each component in CalibSleep, ablation experiments are conducted on the in-house dataset (Tab. 2). Single-stream models and

Table 2. Ablation study of CalibSleep on the No.6 Hospital dataset.

Variant	Time	TF	CMC	Rule	Acc	MF1	Kappa
A: Signal only	✓				0.827	0.815	0.809
B: Time–Freq only		✓			0.814	0.803	0.799
C: Dual-stream (concat)	✓	✓			0.836	0.833	0.828
D: Dual-stream + CMC	✓	✓	✓		0.872	0.846	0.841
E: + CMC + Rule (Full)	✓	✓	✓	✓	**0.896**	**0.877**	**0.874**

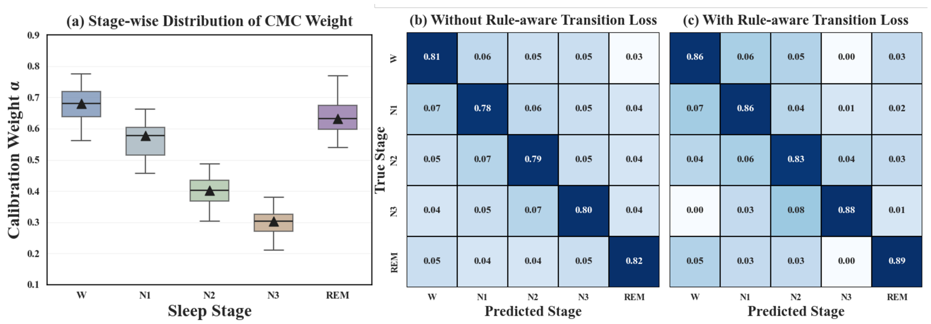

Fig. 3. Interpretability Analysis of CMC and Rule-Aware Classification.

simple feature concatenation yield inferior performance, indicating that neither modality alone nor naive fusion is sufficient. Introducing the proposed cross-modal calibration module improves accuracy from 0.836 to 0.872 and macro-F1 from 0.833 to 0.846. Further incorporating rule-aware stage transition regularization achieves the best performance, with an accuracy of 0.896, macro-F1 of 0.877, and kappa of 0.874, validating the effectiveness of all components.

3.4 Interpretability Analysis

To investigate the interpretability of CalibSleep, we randomly selected 100 subjects from the SHHS test set and analyzed the stage-wise behavior of the calibration weight α. As shown in Fig. 3(a), α exhibits clear sleep-stage-dependent patterns. Higher α values during Wake suggest greater reliance on time-domain EEG/EOG waveform information, consistent with the diagnostic relevance of eye movements and wake-related dynamics. In contrast, lower α values during N1–N3 indicate increased dependence on time–frequency representations, where spectral cues such as sleep spindles and slow-wave activity are informative. The rebound of α during REM further agrees with the importance of EOG-related features for identifying rapid eye movements, suggesting that the calibration module learns physiologically meaningful modality preferences.

Figs. 3(b) and 3(c) compare confusion matrices with and without stage transition regularization. Without transition constraints, the model shows noticeable confusion between physiologically implausible stage pairs, especially N3 and REM. Rule-aware regularization reduces such implausible transitions, leading to more concentrated diagonal distributions and improved consistency. These results suggest that the proposed regularization improves both classification performance and the physiological plausibility of the predicted hypnogram.

4 Conclusion

In this paper, we presented **CalibSleep**, a calibrated dual-stream framework for automatic sleep stage classification. CalibSleep jointly models time-domain EEG/EOG signals and time–frequency representations, and combines cross-modal calibration with rule-aware transition regularization to align heterogeneous features, balance modality contributions, and improve physiological consistency. Experiments on the No.6 Hospital dataset and the external SHHS1 dataset demonstrate robust and competitive performance across different data sources, highlighting the effectiveness of the proposed physiologically guided calibration strategy. Future work will validate CalibSleep on larger multi-center cohorts, diverse PSG configurations, and real-world clinical scoring workflows.
Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant No. 62301330.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. M. Ferrara et al.: How much sleep do we need? In: Sleep Med. pp. 155-179 (2001)
2. K. Wulff, S. Gatti.: Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. In: Nat. Rev. Neurosci. pp. 589-599 (2010)
3. V. Birrer, M. Elgendi, O. Lambercy, C. Menon.: Evaluating reliability in wearable devices for sleep staging. In: NPJ Digit. Med. pp. 74 (2024)
4. C. Iber.: The AASM manual for the scoring of sleep and associated events: Rules Termin. In: Tech. Specif. (2007)
5. A. Malhotra, M. Younes, S.T. Kuna, R. Benca, C.A. Kushida, J. Walsh, A. Hanlon, B. Staley, A.I. Pack, G.W. Pien.: Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. In: Sleep. pp. 573-582 (2013)
6. H. Yue, Z. Chen, W. Guo, L. Sun, Y. Dai, Y. Wang, W. Ma, X. Fan, W. Wen, W. Lei.: Research and application of deep learning-based sleep staging: Data, modeling, validation, and clinical practice. In: Sleep Med. (2024)
7. K.-M. Rytkönen, J. Zitting.: Automated sleep scoring in rats and mice using the naive Bayes classifier. In: J. Neurosci. Methods. pp. 60-64 (2011)
8. O. Tsinalis, P.M. Matthews, Y. Guo.: Automatic sleep stage scoring with single-channel eeg using convolutional neural networks. In: arXiv:1610.01683 (2016)
9. A. Ito and T. Tanaka.: SleepSatelightFTC: A Lightweight and Interpretable Deep Learning Model for Single-Channel EEG-Based Sleep Stage Classification. In: IEEE Access. pp. 46263-46272 (2025)

10. N. Sridhar, et al.: Deep learning for automated sleep staging using instantaneous heart rate. In: Npj Digit. Med. pp. 1-10 (2020)
11. J. Wang, X. Wang, X. Ning, Y. Lin, H. Phan and Z. Jia.: Subject-Adaptation Salient Wave Detection Network for Multimodal Sleep Stage Classification. In: IEEE Journal of Biomedical and Health Informatics. pp. 2172-2184 (2025)
12. Z. Ren, J. Ma and Y. Ding.: FlexibleSleepNet:A Model for Automatic Sleep Stage Classification Based on Multi-Channel Polysomnography. In: IEEE Journal of Biomedical and Health Informatics. pp. 3488-3501 (2025)
13. E. Eldele, Z. Chen, C. Liu, M. Wu, C.-K. Kwoh, X. Li, C. Guan.: An attention-based deep learning approach for sleep stage classification with single-channel EEG. In: IEEE Trans. pp. 809-818 (2021)
14. C. Zhao, J. Li.: SleepContextNet: A temporal context network for automatic sleep staging based single-channel EEG. In: Comput. Methods Programs Biomed. (2022)
15. S. Biswal, H. Sun, B. Goparaju, M. B. Westover, J. Sun, and M. T. Bianchi.: Expert-level sleep scoring with deep neural networks. In: J. Amer. Med. Inform. pp. 1643–1650 (2018)
16. H. Phan, O.Y. Chén, M.C. Tran, P. Koch, A. Mertins, M. De Vos.: XSleepNet: Multi-view sequential model for automatic sleep staging. In: IEEE Trans. Pattern Anal. pp. 5903-5915 (2021)
17. Z. Jia, X. Cai, Z. Jiao.: Multi-modal physiological signals based squeeze-and-excitation network. In: IEEE Sens. pp. 3464-3471 (2022)
18. H. Zan, A. Yildiz.: Local pattern transformation-based convolutional neural network for sleep stage scoring. In: Biomed. Signal Process. (2023)
19. Chung J, Gulcehre C, Cho K H, et al.: Empirical evaluation of gated recurrent neural networks on sequence modeling. In: arXiv preprint. (2014)
20. Hu J, Shen L.: Squeeze - and - Excitation Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp.7132-7141 (2018)
21. Wu H, Hu T et al.: TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In: International Conference on Learning Representations. (2023)
22. Y. Dai.:Multichannelsleepnet: A transformer-based model for automatic sleep stage classification with psg. In: IEEE J. Biomed. Heal. Inform. pp. 4204-4215 (2023)
23. H. Phan, O.Y. Chén.: XSleepNet: Multi-view sequential model for automatic sleep staging. In: IEEE Trans. Pattern Anal. pp. 5903-5915 (2021)
24. A. Zaman, S. Kumar, S. Shatabda.: SleepBoost: A multi-level tree-based ensemble model for automatic sleep stage classification. Med. Biol. pp. 2769-2783 (2024)
25. K. Kontra.: Core-sleep: A multimodal fusion framework for time series robust to imperfect modalities. In: IEEE Trans. Neural Syst. Rehabil. pp. 840-849 (2024)
26. Y. Dai et al.: Multichannelsleepnet: A transformer-based model for automatic sleep stage classification with psg. In: IEEE J. Biomed. pp. 4204-4215 (2023)
27. H. Phan.: SeqSleepNet: End-to-end hierarchical recurrent neural network for automatic sleep staging. In: IEEE Trans. Neural Syst. pp. 400-410 (2019)
28. H. Phan.: Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification. In: IEEE Trans. Biomed. pp. 2456-2467 (2022)
29. Y. Dai.: Multichannelsleepnet: A transformer-based model for automatic sleep stage classification with psg. In: IEEE J. Biomed. pp. 4204-4215 (2023)
30. Yang Cao.: CrossFusionSleepNet: A multimodal deep learning model for automatic sleep stage classification. In: Biomedical Signal Processing and Control. (2026)