Multimodal sleep stage classification method based on multi-channel convolution and neural ordinary differential equations

Guoyu Zuo[1, 2][0000-0002-7624-4728], Zhenyu Yang[1,2,3][0009-0001-4605-0260], Erjun Xiao[3][0009-0002-7344-6516] and TieLing Zhang[*3][0000-0002-5111-9891]

[1] School of Information Science and Technology, Beijing University of Technology, Beijing, 100124, China
[2] Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing, 100124, China
[3] Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, 200031,China
zhangtielin@ion.ac.cn

Abstract. Sleep is a critical state characterized by reduced mental and physical activity, with distinct brain wave patterns observed via electroencephalogram (EEG) and electrooculogram (EOG). Adequate sleep is essential for physiological and mental health, while poor sleep quality can impair attention, immunity, and increase chronic disease risk. Accurate sleep staging is vital for assessing sleep quality, but traditional manual methods are time-consuming and subjective. Existing automatic sleep staging approaches struggle to capture temporal dynamics, multi-scale signal characteristics, and multimodal data correlations, often yielding low accuracy, particularly with imbalanced sleep stage distributions. This study proposes a novel hybrid architecture integrating a multi-scale convolutional feature extractor with a Neural Ordinary Differential Equation (Neural ODE) network to enhance classification accuracy by capturing multi-scale temporal and dynamic features of EEG and EOG signals. A Transformer-based multimodal feature fusion mechanism leverages self-attention to establish cross-modal associations, improving feature integration. Additionally, a focal loss function is introduced to address category imbalance by dynamically adjusting weights, enhancing recognition of minority sleep stages. The proposed method offers significant improvements in automatic sleep staging, with potential for clinical and research applications.

Keywords: Deep learning; EEG signal processing; Neural ordinary differential equations; Multimodal fusion.

1 Introduction

Sleep is a vital physiological process that significantly impacts physical and mental health. It is characterized by distinct stages — Wake, Non-Rapid Eye Movement (NREM) stages (N1, N2, N3), and Rapid Eye Movement (REM) — each associated with unique brain wave patterns observable through electroencephalogram (EEG) and electrooculogram (EOG) signals [1]. Accurate classification of these stages is crucial for evaluating sleep quality, diagnosing sleep disorders such as obstructive sleep apnea, and supporting long-term health monitoring [2]. Traditionally, sleep stage scoring relies on manual analysis of polysomnography (PSG) data by experts, which is time-consuming, labor-intensive, and subject to inter-rater variability [3]. The advent of automated sleep stage classification using deep learning has offered a promising solution to enhance efficiency and objectivity, yet existing methods face specific limitations that hinder their performance.

Current deep learning approaches often fail to effectively capture the multi-scale temporal dynamics and complex non-linear patterns in EEG and EOG signals, leading to reduced classification accuracy [4]. Moreover, these methods frequently employ inadequate fusion techniques, which limit their ability to fully exploit the intrinsic correlations between multimodal signals like EEG and EOG, thus missing critical complementary information [5]. Additionally, the imbalanced nature of sleep stage datasets, where minority classes such as N1 (comprising only 2%-5% of samples) are underrepresented, poses a significant challenge for achieving robust classification across all stages [6]. Addressing these limitations is critical for advancing the practical utility of automated sleep stage classification in clinical and research settings.

To overcome these limitations, this paper proposes a novel deep learning framework for sleep stage classification that integrates multi-scale convolutional neural networks (CNNs), neural ordinary differential equations (Neural ODEs), and a Transformer-based cross-attention mechanism for effective multimodal EEG and EOG fusion. Our approach makes the following key contributions:

1. Multi-Scale CNN and Neural ODE Integration: We introduce a hybrid architecture combining a multi-scale CNN feature extractor with a Neural ODE module. The multi-scale CNN captures temporal features across varying time scales, while the Neural ODE models the continuous dynamic evolution of EEG and EOG signals, enhancing the representation of non-linear temporal patterns.
2. Transformer-Based Cross-Attention Fusion: A cross-attention mechanism is employed to dynamically model and fuse cross-modal dependencies between EEG and EOG signals, enabling complementary feature enhancement and improving classification performance.
3. Focal Loss for Class Imbalance: To address the class imbalance problem, we utilize the focal loss function, which dynamically adjusts sample weights to prioritize minority classes like N1, thereby improving recognition accuracy for underrepresented stages.

2 Related Work

Sleep stage classification has seen significant advancements with the integration of signal processing and deep learning techniques, transitioning from manual feature engineering to automated, data-driven approaches. Traditional methods primarily relied on handcrafted features extracted from electroencephalogram (EEG) signals, such as time-domain statistics, frequency-domain power spectra, or non-linear dynamics, followed by classification using machine learning algorithms like random forests, support vector machines, or decision trees [7, 8]. For example, Hassan et al. [9] employed tunable Q-factor wavelet transform (TQWT) to decompose EEG signals and extract spectral features, achieving a classification accuracy of 91.50% using a random forest classifier on the Sleep-EDF dataset. Liu et al. [2] utilized ensemble empirical mode decomposition (EEMD) to extract statistical and non-linear features from EEG signals, combining them with an XGBoost classifier to achieve 93.1% accuracy. Similarly, Liang et al. [10] leveraged multi-scale entropy and autoregressive features, while Fraiwan et al. [11] used Renyi entropy with random forests, reporting accuracies of 83%. Despite their effectiveness, these methods require extensive feature engineering, which is labor-intensive and often lacks generalizability across diverse datasets or populations.

The advent of deep learning has revolutionized sleep stage classification by enabling end-to-end feature extraction and improving robustness. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), including long short-term memory (LSTM) networks, have become prevalent due to their ability to model the temporal and sequential nature of polysomnography (PSG) signals. Supratak et al. [12] introduced DeepSleepNet, a hybrid model combining CNNs for feature extraction and LSTMs for sequence modeling, achieving 82% accuracy on the Sleep-EDF dataset using single-channel EEG. Mousavi et al. [13] proposed SleepEEGNet, a sequence-to-sequence architecture integrating CNNs and LSTMs, which reported 84.26% accuracy. Phan et al. [14] developed SeqSleepNet, a hierarchical RNN with attention mechanisms, leveraging multi-channel PSG signals to achieve 87.1% accuracy. Tsinalis et al. [15] employed a one-dimensional CNN for end-to-end sleep stage scoring, attaining 82% accuracy, while Ghimatgar et al. [16] combined gated recurrent units (GRUs) with CNNs to reach 89% accuracy by capturing temporal trends.

Beyond EEG, the incorporation of electrooculogram (EOG) signals has shown promise, particularly for distinguishing stages like REM and N1. Fan et al. [17] proposed EOGNet, a deep learning model based on single-channel EOG, achieving 81.2% accuracy on the MASS dataset and demonstrating EOG ' s potential as a standalone or complementary modality. Van Gorp et al. [18] reported 85% accuracy using single-channel EOG for sleep staging in patients with sleep disorders. Estrada et al. [19] highlighted the synergistic role of EOG and electromyogram (EMG) in improving classification accuracy, particularly for complex stages.

Despite these advancements, several challenges persist. First, many models struggle to capture the multi-scale temporal dynamics and continuous non-linear patterns in EEG and EOG signals, limiting their ability to represent complex sleep stage transitions [4]. Second, multimodal fusion is often rudimentary, relying on simple feature concatenation that fails to exploit cross-modal interactions between EEG and EOG [20]. For instance, direct concatenation may introduce redundant or isolated information, reducing discriminative power. Third, class imbalance remains a critical issue, with minority classes like N1 (2%-5% of samples) exhibiting significantly lower classification accuracies (e.g., 30%-40% in some studies [15, 21, 22]). Data augmentation techniques, such as time-shift rolling [23] or SMOTE [24], have been explored to mitigate imbalance, but they risk introducing artifacts or distributional biases. Weighted loss functions, like focal loss [25], have shown promise but are underutilized in sleep staging.

Recent studies have begun exploring advanced techniques to address these gaps. Neural ordinary differential equations (Neural ODEs) [26] offer a continuous modeling framework for temporal dynamics, overcoming the limitations of discrete-layer networks. Transformer-based attention mechanisms [27] have been applied to capture long-range dependencies and cross-modal interactions, though their use in sleep staging is limited. Our work builds on these developments by proposing a novel framework that integrates multi-scale CNNs for robust feature extraction, Neural ODEs for continuous dynamic modeling, Transformer-based cross-attention for multimodal fusion, and focal loss to address class imbalance. This approach achieves superior performance compared to state-of-the-art models, particularly for minority classes, as demonstrated on the Sleep-EDF dataset.

3 Proposed Method

This section presents the proposed deep learning framework for sleep stage classification, designed to address the challenges of multi-scale temporal feature extraction, effective multimodal fusion, and class imbalance in EEG and EOG signals. The framework integrates a multi-scale convolutional neural network (CNN) feature extractor, a Transformer-based cross-attention mechanism for multimodal fusion, a neural ordinary differential equation (Neural ODE) module for continuous dynamic modeling, and a gated recurrent unit (GRU) classifier. The overall architecture is depicted in Figure 1, and each component is detailed below, along with optimization strategies to ensure robust performance.

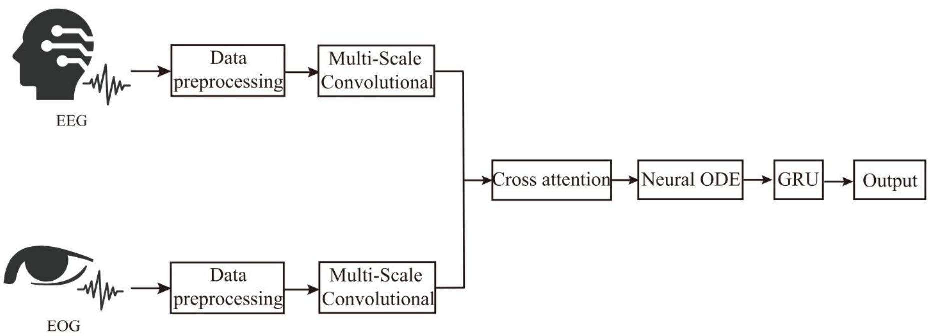

Figure 1: Proposed Model Architecture

3.1 Multi-Scale Convolutional Feature Extractor

EEG and EOG signals exhibit complex temporal patterns across multiple time scales, such as high-frequency sleep spindles (12-14 Hz) in the N2 stage and low-frequency slow-wave activity (0.5-4 Hz) in the N3 stage. To effectively capture these multi-scale characteristics, we design a multi-scale CNN feature extractor with parallel convolutional branches, each tailored to extract features at different temporal resolutions.

For each modality (EEG and EOG), the input signal is divided into 30-second epochs, preprocessed to eliminate noise, and processed through two parallel branches. The short-term branch utilizes a smaller convolutional kernel (size 5) to capture high-frequency, localized patterns, such as sleep spindles or rapid eye movements, followed by a max-pooling layer to reduce dimensionality while retaining key features. In contrast, the medium-term branch employs a larger kernel (size 15) to extract broader temporal patterns, like slow-wave activity, and is similarly followed by a max-pooling layer.

Each convolutional layer is followed by batch normalization to stabilize training and a ReLU activation function to enhance non-linear representation. The convolutional operation for the (i)-th branch is defined as: $y_i = f\left(\sum_j x_j * w_i + b_i\right)$, where x is the input signal, w_i and b_i are the weights and bias, * denotes convolution, and f is the ReLU function. The outputs from both branches are concatenated to form modality-specific feature sequences, X_{EEG} and $X_{EOG} \in R^{T\times d}$, where T is the sequence length and d is the feature dimension. This multi-scale approach ensures comprehensive capture of temporal patterns critical for distinguishing sleep stages.

3.2 Transformer-Based Cross-Attention Fusion

EEG and EOG signals provide complementary information — EEG captures global brain activity patterns, while EOG reflects localized eye movement dynamics. Traditional fusion methods, such as feature concatenation, often fail to exploit cross-modal interactions, leading to suboptimal performance. To address this, we propose a Transformer-based cross-attention mechanism to dynamically model and fuse cross-modal dependencies between EEG and EOG features.

The feature sequences X_{EEG} and X_{EOG} from the multi-scale CNN are mapped to que ry (Q), key (K), and value (V) vectors using learnable weight matrices: $Q_{EEG} = X_{EEG}W_Q$, $K_{EOG} = X_{EOG}W_K$, $V_{EOG} = X_{EOG}W_V$, $Q_{EOG} = X_{EOG}W_Q$, $K_{EEG} = X_{EEG}W_K$, $V_{EEG} = X_{EEG}W_V$. The cross-attention output for EEG is computed as: $\text{Attention}EEG = \text{softmax}\left(\frac{QEEGK_{EOG}^T}{\sqrt{d_k}}\right)V_{EOG}$ and similarly for EOG: $\text{Attention}EOG = \text{softmax}\left(\frac{QEOGK_{EEG}^T}{\sqrt{d_k}}\right)V_{EEG}$. Here, d_k is the dimension of the key vectors, and the softm ax operation ensures normalized attention weights. The attention outputs, representing EEG features enhanced by EOG context and vice versa, are concatenated and passed through additional Transformer layers to produce a fused feature sequence $X_{fused} \in R^{T\times d'}$. This approach captures both modality-specific and cross-modal interactions, en hancing discriminability, particularly for stages like REM, where EOG's rapid eye mo vements correlate with EEG's high-frequency patterns.

3.3 Neural Ordinary Differential Equation

Sleep stage transitions exhibit continuous, non-linear temporal dynamics that are challenging to model with discrete-layer neural networks. To address this, we incorporate a Neural ODE module, which represents feature evolution as a continuous differential equation: $\frac{dh(t)}{dt} = f(h(t), t, \theta)$, where h(t) is the hidden state, f is a neural network parameterized by θ, and t denotes time. The output state is obtained via numerical integration: $h(t_1) = h(t_0) + \int_{t_0}^{t_1} f(h(t), t, \theta)dt$. In our model, the input is the fused feature sequence X_{fused}, and the initial hidden state is initialized as: $h(t_0) = W_{init}X_{fused} + b_{init}$, where W_{init} and b_{init} are learnable parameters. The function f is implemented as a two-layer multilayer perceptron (MLP) with ReLU activation to capture non-linear dynamics. The integration is performed using the Dopri5 Runge-Kutta method for accuracy and efficiency. The output $H \in R^{T\times d''}$ represents an enhanced feature sequence that captures the continuous temporal evolution of the fused EEG and EOG signals, improving the modeling of stage transitions, such as from N1 to N2 or N3 to REM.

3.4 Gated Recurrent Unit Classifier

Sleep stages exhibit sequential dependencies, with transitions following specific patterns (e.g., N1 to N2). To model these dependencies, we employ a GRU as the classifier, which processes the Neural ODE output sequentially. The GRU updates its hidden state using update and reset gates: $z_t = \sigma(W_z[h_{t-1}, x_t] + b_z)$, $r_t = \sigma(W_r[h_{t-1}, x_t] + b_r)$, $\tilde{h}_t = \tanh(W_h[r_t \odot h_t - 1, x_t] + b_h)$, $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$, where $x_t \in H$ is the input at time t, h_t is the hidden state, σ is the sigmoid function, and $\odot$ denotes element-wise multiplication. The final hidden state h_T is passed through a fully connected layer to predict stage probabilities: $\mathrm{P} = \mathrm{softmax}(W_{fc}h_T + b_{fc})$, where $P \in R^C$ and C = 5 (W, N1, N2, N3, REM). The GRU' s hidden dimension is set to 128, with a 10% dropout applied to prevent overfitting. The sequential processing ensures that the model captures stage transition patterns, enhancing classification accuracy.

3.5 Optimization Strategies

To mitigate the impact of class imbalance, particularly for the N1 stage (2%-5% of samples), we use the focal loss function: $\mathrm{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$, where p_t is the predicted probability for the true class, α_t is a class-specific weight higher for minority classes, and $\gamma = 2$ focuses on hard-to-classify samples. This approach prioritizes minority classes, improving their recognition accuracy.

A 10% dropout is applied after the convolutional, cross-attention, and GRU layers to prevent overfitting and enhance generalization, particularly given the high-dimensional EEG and EOG features.

Adam Optimizer with Cosine Annealing: The Adam optimizer is used with an initial learning rate of 10^{-3}. To improve convergence, we employ a cosine annealing scheduler to adjust the learning rate over a cycle of 50 epochs: $\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} -$

$\eta_{min})\left(1+\cos\left(\frac{t}{T}\pi\right)\right)$, $\eta_{max} = 10^{-3}, \eta_{min} = 10^{-5}$, T = 50. This strategy enables rapid exploration in early training and fine-tuned optimization in later stages.

The combination of these components and optimization strategies enables the proposed model to effectively capture multi-scale features, fuse multimodal signals, model continuous dynamics, and handle class imbalance, leading to robust sleep stage classification performance.

4 Experimental Setup and Results

This section describes the experimental setup, evaluation metrics, and results of the proposed deep learning framework for sleep stage classification. The model is evaluated on the Sleep-EDF Database and its expanded version, with comparisons against state-of-the-art baselines and ablation studies to assess the contribution of each component. The experiments aim to validate the effectiveness of the proposed multi-scale convolutional neural network (CNN), Transformer-based cross-attention fusion, neural ordinary differential equation (Neural ODE), and focal loss in addressing multi-scale feature extraction, multimodal fusion, and class imbalance.

4.1 Dataset and Preprocessing

The Sleep-EDF Database and its expanded version, hosted by PhysioNet, are widely used benchmarks for sleep stage classification [28]. The dataset comprises 197 whole-night polysomnography (PSG) recordings, including electroencephalogram (EEG, Fpz-Cz channel), electrooculogram (EOG), electromyogram (EMG), and additional physiological signals, with a sampling frequency of 100 Hz. Sleep stages are annotated according to the Rechtschaffen and Kales standard [29] and converted to the American Academy of Sleep Medicine (AASM) standard, categorizing stages into five classes: Wake (W), Non-Rapid Eye Movement (N1, N2, N3), and Rapid Eye Movement (REM). The expanded dataset includes additional recordings from the Sleep Cassette Study (1987-1991) and Sleep Telemetry Study (1994), totaling 153 and 44 records, respectively [30].

Preprocessing is performed to ensure signal quality and compatibility with the model. EEG and EOG signals are bandpass-filtered to remove noise and artifacts, retaining frequencies relevant to sleep stages (0.5-30 Hz for EEG, 0.05-3.5 Hz for EOG). The signals are segmented into 30-second epochs, each labeled with one of the five sleep stages. Epochs with ambiguous or missing annotations are discarded. The dataset is split into 80% for training and 20% for testing, ensuring a balanced representation of sleep stages across splits.

4.2 Experimental Settings

The proposed model is implemented in PyTorch and trained on a single NVIDIA GPU. Hyperparameters are tuned using grid search and validated on a hold-out validation set. The final model configuration is designed to optimize performance in sleep stage classification. It employs the Adam optimizer with an initial learning rate of 10^{-3}, β_1 = 0.9, and β_2 = 0.999, paired with a cosine annealing learning rate scheduler that cycles every 50 epochs and reduces the learning rate to a minimum of

10^{-5}. Training is conducted with a batch size of 64 over 500 epochs. To address class imbalance, focal loss is used with $\gamma = 2$ and class weights (α_t) set inversely proportional to class frequency, prioritizing minority classes like N1. A 10% dropout is applied after the convolutional, cross-attention, and GRU layers, with the GRU hidden dimension set to 128.

In this paper, performance is assessed using two metrics: Overall Accuracy (ACC), which measures the percentage of correctly classified epochs, and Macro-Averaged F1 Score (Macro-F1), the harmonic mean of precision and recall averaged across all classes, ensuring balanced evaluation despite class imbalance.Results

4.3 Contrast Experimental

The confusion matrix for the full model, shown in Figure 2, illustrates balanced performance across all sleep stages. Notably, the model achieves high accuracy for the N1 stage, which is typically challenging due to its low sample proportion (2%-5%). Misclassifications are minimal, primarily occurring between adjacent stages (e.g., N1 and N2), reflecting the model's ability to capture subtle stage transitions. The cross-attention mechanism and focal loss contribute significantly to improved N1 and REM classification compared to baselines.

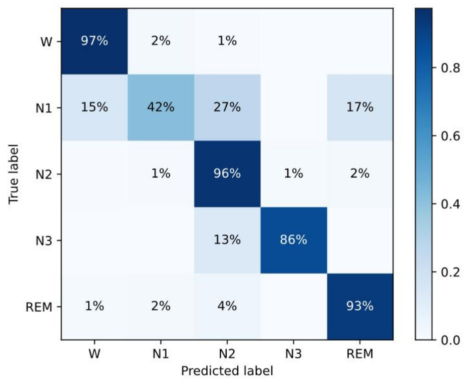

Figure 2: Confusion Matrix of the Proposed Model

Figure 3 plots the test accuracy over training epochs, showing stable convergence due to the cosine annealing scheduler and Adam optimizer. The model reaches peak performance around 400 epochs, with no signs of overfitting, as evidenced by consistent validation performance (not shown for brevity).

The proposed model is compared with four state-of-the-art baseline models: DeepSleepNet [10], SleepEEGNet [11], SeqSleepNet [12], and EOGNet [13]. These models represent diverse approaches, including CNN-LSTM hybrids, sequence-to-sequence architectures, hierarchical RNNs, and EOG-based methods. The results, reported as mean ACC and Macro-F1 on the Sleep-EDF test set, are shown in Table 1.

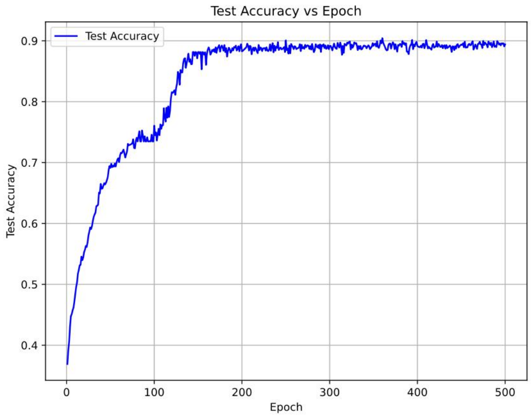

Figure 3: Test Accuracy vs. Training Epochs

Table 1. Comparative Results on Sleep-EDF Dataset

Model	ACC(%)	Macro-F1 (%)
Proposed Model	89.40	84.22
DeepSleepNet	82.00	76.90
SleepEEGNet	84.26	79.66
SeqSleepNet	87.10	83.30
EOGNet	76.30	69.30

The proposed model achieves the highest performance, with an ACC of 89.40% and a Macro-F1 of 84.22%, outperforming all baselines. Compared to DeepSleepNet (82.00% ACC), the proposed model benefits from multi-scale feature extraction and continuous dynamic modeling via Neural ODE. SleepEEGNet (84.26% ACC) and SeqSleepNet (87.10% ACC) perform well but are limited by simpler fusion strategies and discrete-layer architectures. EOGNet, relying solely on EOG signals, yields the lowest performance (76.30% ACC), underscoring the importance of multimodal EEG-EOG fusion. The high Macro-F1 score of the proposed model indicates robust performance across all classes, particularly for minority classes like N1, attributed to the focal loss and cross-attention mechanism.

4.4 Ablation Study

To evaluate the contribution of each component, we conduct ablation studies by removing or replacing key modules. The configurations tested are:

- w/o Neural ODE: Replacing the Neural ODE with an MLP results in a significant performance drop (ACC: 80.02%, Macro-F1: 72.10%), particularly for N3 and REM stages, where continuous temporal dynamics are critical. This highlights the Neural ODE ' s role in modeling non-linear stage transitions.

- w/o Cross-Attention: Feature concatenation leads to a 15.59% drop in ACC (73.81%) and a 15.96% drop in Macro-F1 (68.26%), with notable degradation in N1 and REM classification. This underscores the importance of dynamic cross-modal interactions for effective EEG-EOG fusion.
- w/o Multi-Scale CNN: Using a single-scale CNN reduces ACC to 70.13% and Macro-F1 to 62.73%, with increased errors in N1 and REM stages. The multi-scale CNN is essential for capturing diverse temporal patterns like sleep spindles and slow-wave activity.
- w/o Focal Loss: Cross-entropy loss yields an ACC of 83.79% and a Macro-F1 of 75.90%, with a significant decline in N1 classification performance. The focal loss effectively addresses class imbalance, improving minority class recognition.

The ablation study confirms that each component — multi-scale CNN, cross-attention, Neural ODE, and focal loss — plays a critical role in achieving high performance, with synergistic effects enhancing overall classification accuracy and robustness.

Table 2: Ablation Study Results on Sleep-EDF Dataset

Configuration	ACC(%)	Macro-F1 (%)
Full Model	89.40	84.22
w/o Neural ODE (replaced with MLP)	80.02	72.10
w/o Cross-Attention (concatenation)	73.81	68.26
w/o Multi-Scale CNN (single-scale)	70.13	62.73
w/o Focal Loss (cross-entropy)	83.79	75.90

The confusion matrix of the ablation experiment is shown in Figure 4. Figure 4 (a) shows the experimental result of replacing Neural ODE with MLP, Figure 4 (b) shows the experimental result of replacing cross attention with feature concatenation, Figure 4 (c) shows the experimental result of replacing multi-scale convolution with single-scale convolution, and Figure 4 (d) shows the experimental result of replacing focal loss with cross entropy loss.

4.5 Discussion

The proposed model excels in sleep stage classification by effectively tackling several critical challenges. Its multi-scale CNN captures diverse temporal patterns, enhancing the classification of stages with distinct frequency characteristics, such as N2 spindles and N3 slow waves. Additionally, the Transformer-based cross-attention mechanism facilitates multimodal fusion, improving EEG-EOG synergy, particularly for REM and N1 stages, where eye movements and brain activity are closely correlated. The model also employs Neural ODEs to capture continuous, non-linear stage transitions, outperforming discrete-layer alternatives like MLPs. Furthermore, the use of focal loss addresses class imbalance, significantly improving performance on minority classes and achieving robust Macro-F1 scores.

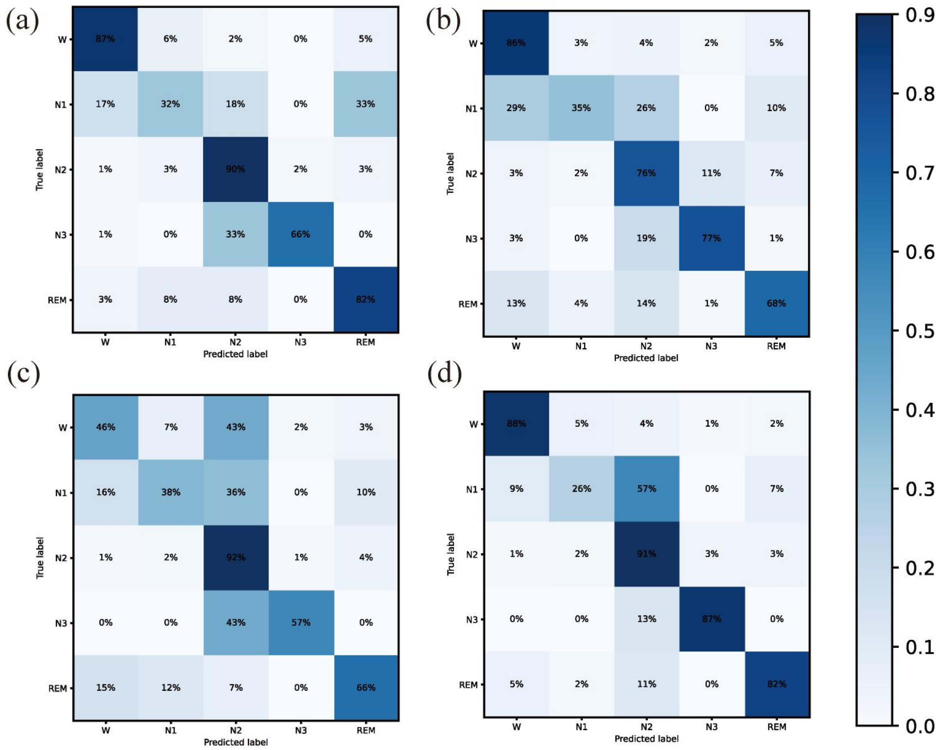

Figure 4: The confusion matrix of the ablation experiment

However, limitations exist. The computational complexity of the cross-attention mechanism and Neural ODE's numerical solver may hinder deployment on resource-constrained devices, such as wearable sleep monitors. Additionally, signal noise and artifacts can affect Neural ODE stability, necessitating robust preprocessing.

5 Conclusion

This paper presents a novel deep learning framework for automated sleep stage classification, achieving an accuracy of 89.40% and a macro-averaged F1 score of 84.22% on the Sleep-EDF Database, surpassing baselines like DeepSleepNet (82.00% ACC), SleepEEGNet (84.26% ACC), SeqSleepNet (87.10% ACC), and EOGNet (76.30% ACC). The framework integrates a multi-scale CNN for temporal feature extraction, a Transformer-based cross-attention mechanism for EEG-EOG fusion, a Neural ODE for continuous temporal modeling, and a GRU classifier with focal loss to address class imbalance, with ablation studies confirming each component's critical role (e.g., 19.27% ACC drop without multi-scale CNN). Notably, the model excels in classifying minority classes like N1, offering significant potential for clinical applications such as automated sleep disorder diagnosis and real-time monitoring

Acknowledgement

This work was supported by the Beijing Nova Program (No. 20230484369), the Youth Innovation Promotion Association CAS, and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB1010302).

References

1. Liu, P., Qian, W., Zhang, H., et al.: Automatic sleep stage classification using deep learning: signals, data representation, and neural networks, 57, 301 (2024).
2. Liu, C., Tan, B., Fu, M., et al.: Automatic sleep staging with a single-channel EEG based on ensemble empirical mode decomposition, 567, 125685 (2020).
3. Rundo, J. V., Downey III, R.: Polysomnography, 160, 381-392 (2019).
4. Phan, H., Mikkelsen, K.: Automatic sleep staging of EEG signals: recent development, challenges, and future directions, 43, 04TR01 (2022).
5. Penzel, T.: Sleep scoring moving from visual scoring towards automated scoring, 45, zsac190 (2022).
6. Patanaik, A., Ong, J. L., Gooley, J. J., et al.: An end-to-end framework for real-time automatic sleep stage classification, 41, zsy041 (2018).
7. Biswal, S., Sun, H., Goparaju, B., et al.: Expert-level sleep scoring with deep neural networks, 25, 1643-1650 (2018).
8. Ebrahimi, F., Setarehdan, S. K., Ayala-Moyeda, J., et al.: Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals, 112, 47-57 (2013).
9. Wolpert, E. A.: A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects, 20, 246 (1969).
10. Hassan, A. R., Bhuiyan, M. I. H.: A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features, 271, 107-118 (2016).
11. Liang, S. F., Kuo, C. E., Hu, Y. H., et al.: Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models, 61, 1649-1657 (2012).
12. Sen, B., Peker, M., Çavuşoğlu, A., et al.: A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms, 38, 1-21 (2014).
13. Fraiwan, L., Lweesy, K., Khasawneh, N., et al.: Automated sleep stage identification system based on time – frequency analysis of a single EEG channel and random forest classifier, 108, 10-19 (2012).
14. Zhuang, L., Dai, M., Zhou, Y., et al.: Intelligent automatic sleep staging model based on CNN and LSTM, 10, 946833 (2022).
15. Sridhar, N., Shoeb, A., Stephens, P., et al.: Deep learning for automated sleep staging using instantaneous heart rate, 3, 106 (2020).
16. Tsinalis, O., Matthews, P. M., Guo, Y., et al.: Automatic sleep stage scoring with single-channel EEG using convolutional neural networks, 1610, 01683 (2016).
17. Ghimatgar, H., Kazemi, K., Helfroush, M. S., et al.: Neonatal EEG sleep stage classification based on deep learning and HMM, 17, 036031 (2020).
18. Supratak, A., Dong, H., Wu, C., et al.: DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG, 25, 1998-2008 (2017).

19. Mousavi, S., Afghah, F., Acharya, U. R.: SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach, 14, e0216456 (2019).
20. Phan, H., Andreotti, F., Cooray, N., et al.: SeqSleepNet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging, 27, 400-410 (2019).
21. Fan, J., Sun, C., Chen, C., et al.: EEG data augmentation: towards class imbalance problem in sleep staging tasks, 17, 056017 (2020).
22. Tsinalis, O., Matthews, P. M., Guo, Y.: Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders, 44, 1587-1597 (2016).
23. Khalili, E., Asl, B. M.: Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel EEG, 204, 106063 (2021).
24. Lashgari, E., Liang, D., Maoz, U.: Data augmentation for deep-learning-based electroencephalography, 346, 108885 (2020).
25. van Gorp, H., van Gilst, M. M., Overeem, S., et al.: Single-channel EOG sleep staging on a heterogeneous cohort of subjects with sleep disorders, 45, 055007 (2024).
26. Fan, J., Sun, C., Long, M., et al.: EOGnet: A novel deep learning model for sleep stage classification based on single-channel EOG signal, 15, 573194 (2021).
27. Estrada, E., Nazeran, H., Barragan, J., et al.: EOG and EMG: Two important switches in automatic sleep stage classification, 2006, 2458-2461 (2006).
28. Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A. C., Oberyé, J. J. L.: Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG, 47, 1185-1194 (2000).
29. Rechtschaffen, A., Kales, A.: A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects, Not Applicable, Not Applicable (1968).
30. Tian, Y., Liu, G., Taj, I. A., Sun, Z., Zhou, Z., Zheng, W.: Removal of Artifacts from EEG Signals: A Review, 19, 987 (2019).
31. Jasper, H. H.: The 10-20 electrode system of the International Federation, 10, 371-375 (1958).