Cross-LUSleepNet: A U-shaped Sleep Staging Method Based on Cross-Layer Connection Modules and Bi-LSTM

Yulin Gong, Jinrui Zhang, Yudan Lv, Chang Liu and Xiaojuan Chen

Abstract—**Despite the development of numerous sleep staging algorithms, their application in clinical environments remains limited. This is primarily due to significant differences in EEG signals between patients with sleep disorders and healthy individuals, caused by diverse pathological factors and substantial inter-individual variability. To address these challenges, we propose a U-shaped neural network with cross-layer connections combined with a Bidirectional Long Short-Term Memory (Bi-LSTM) model for automatic sleep staging. Specifically, we design a cross-layer connection module to integrate features from adjacent layers and incorporate them into the skip connections of the U-shaped architecture. Additionally, a Bi-LSTM module is embedded between specific feature extraction and fusion modules to enhance the continuity of global features and the representation of contextual information. To validate the effectiveness of our approach, we conducted experiments on 44 patients with various sleep disorder pathologies and evaluated the model on a public dataset. The results demonstrate that our model significantly improves sleep staging accuracy in clinical patient populations.**

Index Terms—**EEG signal, Sleep Stage Classification, Bi-LSTM, Deep Neural Networks**

I. Introduction

SLEEP is essential to human health, serving not only as a foundation for physical recovery but also playing critical roles in emotional regulation, immune function, and cognitive support. Studies have shown that sleep disorders can lead to physiological dysfunction and are important indicators for the early detection of various neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease [1]. Accurate classification of sleep stages is crucial for the diagnosis and treatment of sleep-related disorders. Sleep is a complex and dynamic process, and its precise staging aids in managing conditions like insomnia and sleep apnea. Traditionally, sleep staging relies on polysomnography (PSG), which collects multiple physiological signals—such as electrocardiogram (ECG), electromyogram (EMG), electroencephalogram (EEG), and electrooculogram (EOG)—to monitor bodily functions during sleep [2].

According to the standards of the American Academy of Sleep Medicine (AASM), sleep is categorized into wakefulness (W), rapid eye movement (REM), and non-rapid eye movement (NREM). NREM is further divided into three stages: N1 (light sleep), N2 (intermediate sleep), and N3 (deep sleep) [3]. As the gold standard, PSG typically segments an entire night of sleep into 30-second epochs, which are then manually annotated by experienced sleep experts. However, manual sleep staging is tedious, time-consuming, and subjective, with a high probability of human error [4]. As a result, the development of automated sleep staging methods has garnered increasing interest in recent years.

Traditional machine learning-based sleep staging approaches rely on handcrafted features and classifiers such as random forests [5] and hidden Markov models [6]. These methods suffer from subjectivity, limited scalability, and difficulty in handling high-dimensional EEG data, limiting their ability to extract rich features from such complex signals.

With the rapid advancement of deep learning technologies, various neural network models have been proposed for automated sleep staging using physiological time-series data. These models automatically extract hierarchical features, significantly improving classification performance [7]. Early work using stacked sparse autoencoders (SAEs) laid the foundation for subsequent applications of recurrent neural networks (RNNs) [8]. Michielli et al. [9] proposed an RNN architecture based on dual long short-term memory (LSTM) blocks that, after dimensionality reduction of 55 time- and frequency-domain features, significantly improved performance, especially for the challenging N1 stage. Zhang et al.[10]transformed EEG signals into video-like representations, then applied a variant of convolutional neural networks (CNNs)—orthogonal CNN (OCNN)—to overcome limitations of conventional CNNs, achieving superior results. L-SeqSleepNet [11], a sequence-to-sequence model, enhanced sleep staging performance by capturing temporal dependencies more effectively. More recent models have incorporated attention mechanisms and Transformer-based architectures [12]–[16], which improve model interpretability

Manuscript received April 11, 2025; revised September 21, 2025.

This work was supported in part by Jilin Province Science and Technology Development Plan Project (20240304052SF).

Yulin Gong is a professor from School of Electronic Information, Changchun University of Science and Technology, Changchun 130022, China (e-mail: gongyulin@cust.edu.cn).

Jinrui Zhang is a postgraduate student of School of Electronic Information, Changchun University of Science and Technology, Changchun 130022, China (e-mail: 17643194196@163.com).

Yudan Lv is a professor from School of The Department of Neurology, Jilin University, Changchun 130012, China (e-mail: lvyudan@sina.com).

Chang Liu is an associate professor of The Department of Neurology, Jilin University, Changchun 130012, China (e-mail: drliuchang@jlu.edu.cn).

Xiaojuan Chen is a professor from School of Electronic Information, Changchun University of Science and Technology, Changchun 130022, China (e-mail: cxj_neiep@126.com).

by highlighting key time steps or channels [17], thereby enhancing the model's ability to capture salient features and contextual dependencies in sleep data.

Despite these advancements, many existing methods still struggle to capture long-range dependencies in EEG signals. With continued progress in computing, deep learning has become the mainstream approach for sleep staging, improving accuracy and efficiency [18].

Moreover, studies by the International Sleep Association have highlighted differences between healthy individuals and those with sleep disorders. While healthy subjects typically experience a cyclical progression through different sleep stages, patients with sleep disorders often display irregular stage transitions. For example, individuals with sleep apnea tend to have more frequent arousals and reduced proportions of deep sleep. Such irregularities can degrade the performance of sleep staging models and impact the assessment of sleep quality.

In addition, many studies utilize multi-modal input signals to improve staging performance. However, acquiring data from multiple channels can disrupt the subject's sleep, thus potentially affecting both the quality of the sleep itself and the accuracy of the staging results.

II. Related work

Automatic sleep staging has made significant progress in recent years, with classification algorithms generally falling into two categories: traditional machine learning models and deep learning methods. Traditional machine learning [19] relies on manually selecting features from EEG signals and then classifying them. Significant progress has been made in automatic sleep staging in recent years, with existing classification algorithms generally falling into two categories: traditional machine learning approaches and deep learning methods.

Traditional machine learning methods [19] typically rely on manually extracting features from EEG signals, followed by classification using standard models such as support vector machines (SVMs) [20] and k-nearest neighbors (KNN) [21]. These models aim to learn the mapping between the extracted feature space and the corresponding sleep stages. The effectiveness of such methods heavily depends on the quality of the handcrafted features and the optimization of classifier parameters. However, the feature engineering process is often time-consuming and labor-intensive, which not only complicates the sleep staging process but also increases computational and human costs.

With the remarkable success of deep learning in fields such as image recognition and natural language processing, researchers have increasingly explored its potential for processing physiological signals. For instance, SleepEEGNet [22] proposed a model for single-channel EEG sleep staging that extracts time-invariant and frequency-domain features from raw signals. By incorporating a novel loss function, the model reduces the adverse effects of class imbalance during training. DeepSleepNet [23] introduced a two-step training strategy that models the transition rules between sleep stages, improving training efficiency and model stability, while enhancing adaptability to diverse datasets. Building upon this, Akara and Yike [24] developed TinySleepNet, a lightweight CNN-based model designed for resource-constrained environments. By significantly reducing the number of parameters in the feature extraction layers, the model supports faster training and inference with minimal performance loss. XSleepNet [25] addresses the limitations of single-view learning by simultaneously processing both raw EEG signals and their time–frequency representations. By leveraging complementary information from multiple views, the model more effectively captures the underlying data distribution. SleepUTime [26] employs a fully feedforward deep neural network to segment physiological time series and map inputs of arbitrary length to sleep stage sequences across flexible time scales.

While these models have demonstrated strong performance in automatic sleep staging, they often fall short when applied to clinical EEG data. Specifically, they struggle to capture significant waveform patterns and spatial relationships across EEG channels, which are crucial in pathological cases. Moreover, clinical patients often exhibit irregular sleep stage transitions, and current deep learning models are not well-suited to accurately identify and model these irregularities.

To address these challenges, Jia et al. [27] proposed GraphSleepNet, which leverages graph convolutional networks (GCNs) for improved brain connectivity and activity representation. By integrating spatial and temporal attention mechanisms, the model captures both inter-channel spatial relationships and temporal dynamics across adjacent time windows. More recently, SwinSleep [28], an adaptation of the Swin Transformer architecture, has been developed specifically for clinical PSG data. By effectively modeling spatiotemporal dependencies, SwinSleep enhances sleep staging performance in complex clinical scenarios.

III. Method

In this study, we propose a novel sleep staging algorithm named CrossLUSleepNet, which integrates an improved U-Net architecture with a bidirectional long short-term memory network (Bi-LSTM) to address the limited generalization ability of existing sleep staging methods. The overall framework of the proposed method is illustrated in Fig. 01.

The proposed architecture incorporates Bi-LSTM modules into an enhanced U-shaped network, enabling the capture of both fine-grained details and global contextual information in EEG signals through multi-level, continuous convolutional layers. This multi-scale feature extraction strategy enhances the model's ability to identify sleep-specific waveforms across different stages. Furthermore, we introduce a crosslayer connection module designed to reinforce contextual information flow, allowing the model to better represent temporal continuity across sleep transitions. This design significantly improves the model's robustness and accuracy, especially under varying signal-to-noise ratios, while maintaining the lightweight and generalizable characteristics of U-Net and Bi-LSTM architectures.

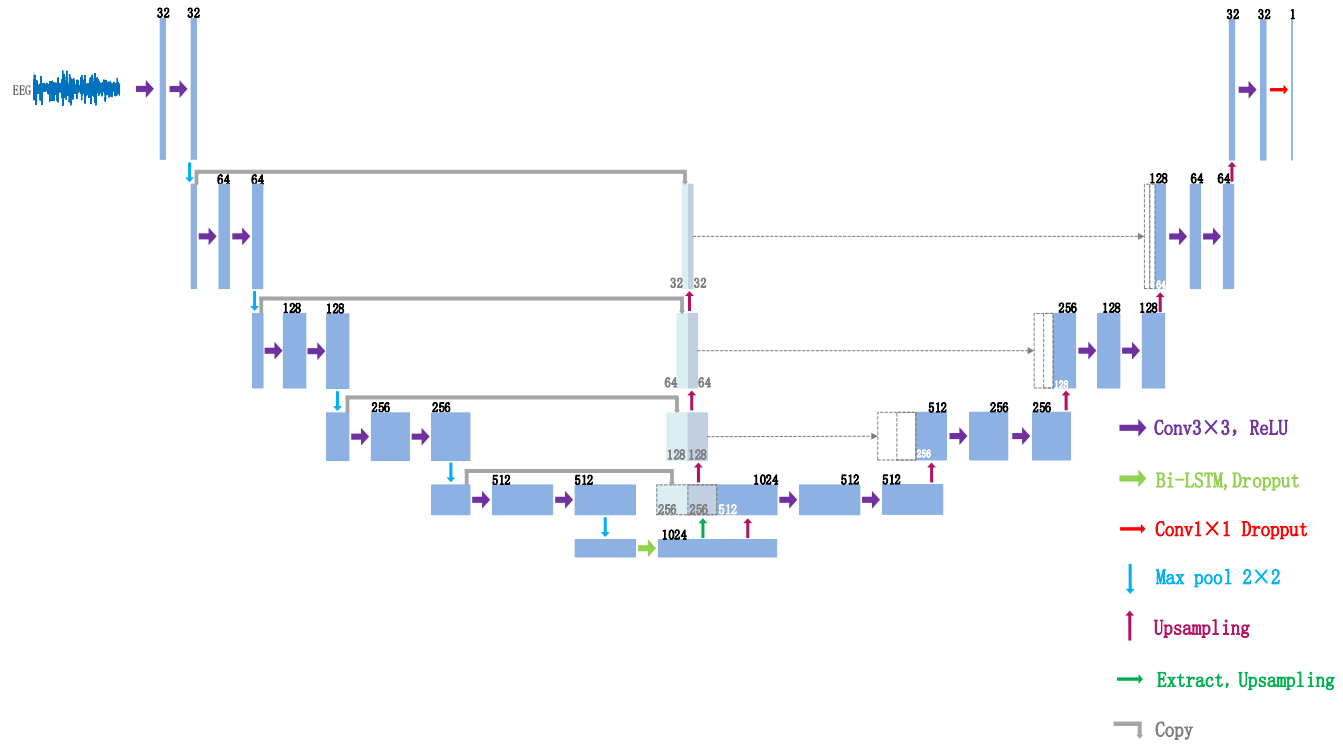


Fig. 1. The overall architecture of Cross-LUSleepNet. It consists of a U-shaped network structure and a cross-layer connection module. The U-shaped network structure consists of an encoder, a decoder, and a Bi-LSTM module. The cross-layer connection module extracts feature information and passes it to adjacent layers for multi-scale feature fusion. The U-shaped network structure abstractly represents multi-scale features through the encoder and decoder combined with the Bi-LSTM module to generate feature maps for different sleep stages.

A. U-shaped network structure

U-Net is a fully convolutional neural network [29] that uses a symmetrical encoder-decoder structure and can effectively capture multi-scale features. The encoder extracts high-level features of the input signal, and the decoder gradually restores the spatial resolution. This mechanism is very helpful for processing complex patterns in sleep EEG signals [30]. Secondly, the skip connection in U-Net allows high-resolution features to be passed directly from the encoder to the decoder, which can preserve more details. In addition, U-Net performs well in tasks such as image segmentation [31], and the sleep staging task can be analogized to the "segmentation" of time series, that is, dividing long-term sleep data into different stages, so the characteristics of U-Net are very suitable for this task. At the same time, the scalability of U-Net allows it to be combined with other models, such as Bi-LSTM [32], to further improve the ability to capture long-term dependencies, thereby enhancing the accuracy of sleep staging.

The encoder utilized in this study comprises five convolutional blocks, each designed to preserve input dimensions through zero padding. Each block contains two consecutive convolutional operations with a 3×3 kernel, followed by batch normalization and a 2×2 max pooling operation with a stride of 2, facilitating downsampling. At each downsampling step, the number of feature channels doubles, for a total of five downsampling steps. This progressive downsampling decreases the input dimensions by a factor of 10 at the lowest level, significantly reducing computational and memory requirements. By downsampling features, the model learns abstract representations at deeper levels, while multi-scale stacked convolutions provide an expanded receptive field in the encoder's final convolutional layer. Subsequently, the outputs from two LSTM layers are progressively aggregated through a Bi-LSTM layer with 1024 units, followed by a Dropout layer to enhance generalization before entering the U-network's expansion path.

The decoder is composed of five convolutional blocks. Each block in the decoder receives as input the upsampled output from the previous layer, along with the output from the cross-layer connection module. It then performs two convolution operations with a 3×3 kernel, followed by rectified linear unit (ReLU) activation and batch normalization. During decoding, the process reconstructs the original image details from the abstract feature map, connecting the generated feature map with both the corresponding feature map calculated by the encoder at the same scale and the feature map from adjacent layers. This allows the model to simultaneously capture local details and global contextual information, enhancing the integration of multi-scale features. Following five upsampling steps, the model includes a 1×1 convolution layer and a Dropout layer to map the final abstract feature to the filter, applying the Softmax function for final sleep stage classification.

B. Bi-LSTM module

EEG signal variations are influenced by both past and future brain states. Traditional unidirectional recurrent neural networks (RNNs) are inherently limited in their ability to capture bidirectional contextual information within EEG feature sequences. To overcome this limitation, our study employs a Bidirectional Long Short-Term Memory (Bi-LSTM) network to model temporal dependencies in both directions.

LSTM networks address the vanishing gradient problem in standard RNNs by introducing three gating mechanisms—

input, forget, and output gates—which effectively manage long-term dependencies in sequential data. However, conventional LSTMs process data in a single temporal direction (typically forward in time), which restricts their ability to incorporate information from future states.

In contrast, the Bi-LSTM architecture consists of two parallel LSTM layers: one processes the input sequence in the forward direction, while the other processes it in the reverse direction. By concatenating the outputs from both directions at each time step, Bi-LSTM captures richer temporal dependencies and provides a more comprehensive representation of EEG sequences. This bidirectional approach enhances the model's capacity to understand transitions between sleep stages, especially when such transitions exhibit subtle temporal cues from both preceding and succeeding intervals. An illustration of the Bi-LSTM processing flow is shown in Fig. 2.

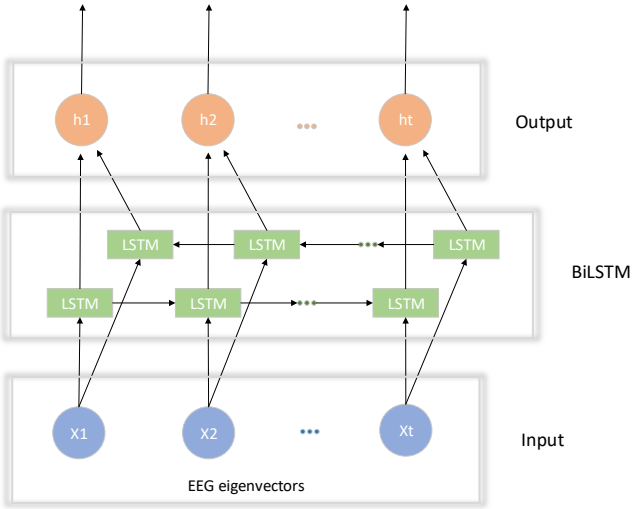


Fig. 2. BiLSTM flow chart.

The principle is as follows:

Among them, $\vec{h}_t$ represents the forward hidden layer state at the moment, $\overleftarrow{h}_t$represents the backward hidden layer state, w_y represents the weight matrix, and b_y represents the bias term.

$$\begin{cases} \vec{h}_t = LSTM(x_t,\ \vec{h}_{t-1}) \\ \overleftarrow{h}_t = LSTM(x_t,\ \overleftarrow{h}_{t-1}) \\ y_t = \sigma\left(w_y \cdot \left|\vec{h}_t, \overleftarrow{h}_t\right| + b_y\right) \end{cases} \quad (1)$$

C. Cross-layer connection module

Neural networks extract hierarchical features from raw data by passing it through multiple convolutional and pooling layers, resulting in multilevel feature maps. In EEG signals, some critical features—such as subtle waveforms or transient events—occur at significantly smaller scales relative to the overall signal. Therefore, a model must effectively integrate information across layers to accurately classify EEG signals at various sleep stages.

However, relying solely on simple upsampling operations in the decoder path can result in substantial information loss, especially for key signal points. To address this, modern deep learning architectures often incorporate mechanisms such as residual connections**,** dense connections**,** and attention modules to facilitate information flow and preserve feature integrity [33].

In this study, we introduce a cross-layer connection module within the skip connections of the U-Net architecture to enhance information fusion across different feature hierarchies. As illustrated in Fig.1, the module comprises four layers and takes two inputs: The downsampled feature map from the previous encoder layer (after max pooling), and The output from the preceding layer of the context information storage module.

Initially, the features output by the LSTM module undergo additional convolution-based feature extraction and are then upsampled using a deconvolution operation with a 2×2 kernel to match the spatial dimensions and channel sizes of the target layer. The resulting feature maps are fused through a combination of element-wise addition and channel-wise concatenation, effectively integrating semantic and contextual information from different network depths.

Finally, these enriched cross-layer features are merged with the upsampled decoder outputs at the corresponding layer. This design facilitates direct information transfer between layers, enabling lower-level features to be propagated rapidly to deeper levels of the network. It compensates for the degradation of low-dimensional details during upsampling and enhances the network's ability to preserve multi-scale and multi-level information.

During training, the cross-layer connection module accelerates the learning of meaningful features, strengthens contextual representation, and improves both the convergence speed and segmentation accuracy of the network.

D. Data augmentation training model

Sleep stage data is usually highly unbalanced, with less data in deep sleep and more data in light sleep and wakefulness. Fig 3 shows clinical sleep data. Since samples on the boundary are often the most difficult to classify, especially when the characteristics of some sleep stages are similar to other stages (such as the boundary between REM and light sleep), N3 uses the borderline oversampling method of Borderline SMOTE to perform data enhancement processing on physiological signal data to overcome the imbalance of sample data. It can effectively help the model better identify different sleep stages and improve the accuracy of the model.

The stages N3 and REM with less data are defined as minority classes, and the stage N2 with more data is defined as majority classes. The 1/2 boundary is used to distinguish safe samples from dangerous samples. If the samples occupy more than half of the k nearest neighbor samples, they are marked as safe samples. Such samples are far from the decision boundary and are easier to be correctly classified, so there is no need to synthesize new samples; the proportion of majority class samples in the k nearest neighbor samples is close to or slightly higher than that of minority class samples, which means that the sample is near the decision boundary. Dangerous samples are the focus of generating new samples, so samples with more than 1/2 belonging to the majority class are defined as dangerous samples. At the same time, all neighbors are majority class samples, indicating that the

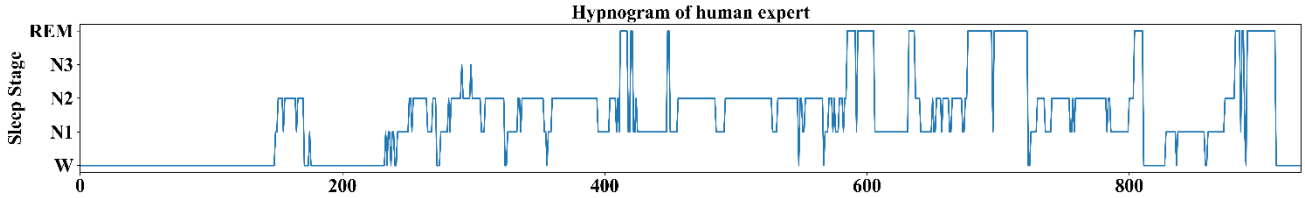


Fig. 3. Clinical sleep staging chart

sample is very likely to be a misclassified point, located in the decision area of the majority class, and is a noise sample, so the samples with all k nearest neighbor samples belonging to the majority class are defined as noise samples. Noise samples are very likely to be in the decision area of the majority class and need to be removed.

The minority class sample M(i), the K nearest neighbor algorithm is used to calculate the k nearest neighbor samples from the entire data set. The distance calculation formula is:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (2)$$

Among them, x_i and y_i are two sample points in ndimensional space, and $d(X,Y)$ is the Euclidean distance between the two sample points.

For each dangerous sample, several minority class samples are randomly selected from its neighborhood. New samples are synthesized by interpolation between these neighbor samples and dangerous samples, thereby expanding the number of minority class samples. Specifically, the formula for generating new samples is:

$$d_{ni} = d_i + rand(0,1) \times (d_{mi} - d_i) \qquad (3)$$

Among them, d_i is a dangerous sample, d_{mi} is a minority class sample in its neighborhood, and $rand(0,1)$ is a random number between 0 and 1. The new samples generated by random interpolation can enrich the minority class samples in the boundary area. The generated new samples are incorporated into the original dataset to form a balanced dataset, so that the model can more fully learn the characteristics of the minority class during training.

IV. Experiments

A. Experimental Data

This study adhered to the principles set forth in the Declaration of Helsinki and was approved by the Institutional Review Board of the First Hospital of Jilin University. Patient medical data were included in the study and used. The dataset contains full-night sleep records of 44 patients with sleep disorders, including 28 male patients and 16 female patients. The dataset contains 8-lead electroencephalogram (EEG) from E1-M2, E2-M1, F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1 channels, 2-lead electrooculogram (EOG), 1-lead mandibular electromyogram (EMG), and manual annotations of various sleep stages. The sampling frequency of EEG and EOG signals was 256Hz. At the same time, a public sleep dataset (Sleep-EDF-78) from Physionet was used for verification [34]. In our experiment, we used the Sleep Box (SC) dataset of sleep-EDF 2018, which contains 153 full-night PSG records from 78 healthy people. Each recording includes two bipolar EEG channels (Fpz-Cz and Pz-Oz), an EOG signal, and a mandibular EMG signal, as well as manual annotations of sleep stages. The EEG and EOG signals were sampled at 100 Hz, and the EMG data was sampled at 1 Hz. The EEG channel signal of Fpz - Cz was used in this study.

TABLE I
PATIENT DEMOGRAPHICS AND CHARACTERISTICS

	Training	Validation	Testing
Number of participants	31	7	6
Age	50 ± 7	52 ± 4	48 ± 6
BMI (kg/m2)	24.8 ± 3.6	25.3 ± 3.4	26.0 ± 4.2
AHI (events/h)	17 ± 3	19 ± 6	10 ± 3
OSA (AHI≥15:<15)	22:9	5:2	1:5
Sleep stage (epoch)	42917	4798	3355
W (%)	10986 (25.6%)	830(17.3%)	439 (13.0%)
N1 (%)	4849(11.3%)	604 (12.6%)	422 (12.5%)
N2 (%)	16737(39.0%)	2097 (43.7%)	1609 (47.9%)
N3 (%)	4206 (9.8%)	412 (8.6%)	218 (6.4%)
R (%)	6137 (14.3%)	854 (17.8%)	667 (19.8%)

B. Data preprocessing

PSG signals, which are closely associated with sleep, primarily consist of electroencephalogram (EEG), electrooculogram (EOG), and electromyography (EMG). Sleep specialists use these three types of signals to identify the characteristic waves of each sleep stage, enabling them to manually label the stages of sleep. For our sleep staging study, we therefore selected a combination of EEG and EOG signals.

We use the same preprocessing steps for all datasets and models. Specifically, the EEG signals are bandpass filtered at 0.5-45 Hz to remove high-frequency noise such as power frequency (50 Hz or 60 Hz), while reducing low-frequency interference such as electromyography. Since EEG signals are easily interfered by artifacts such as eye movements, muscle activity, and heartbeats, independent component

analysis (ICA) is used to separate and remove artifact components, and then the long EEG signals are divided into small segments of 30 seconds. Each segment of data is labeled with the corresponding sleep stage for classification by the model. Finally, the sampling frequency of all signals is unified to 100 Hz.

C. Experimental parameter settings

In the experiments of this study, the network model was built on Python 3.7 with Pytorch 1.12 as the backend deep learning library, trained using the Adam optimizer, $\eta = 1 \times 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1 \times 10^{-8}$, L2 weight regularization with a factor of $1 \times 10^{-6}$was used to prevent overfitting, the batch size was 64, 200 rounds of training were performed, and the model learning rate was 0.001. A tester-independent 10-fold cross-validation was performed on the sleep patient database to fully evaluate the cross-subject performance of the network model. The network model was run on an NVIDIA GTX 3090 Ti GPU.

D. Evaluation Metrics

When evaluating the performance of neural network models in sleep staging tasks, indicators such as precision, recall, and F1 score are usually used to measure the classification effect of each sleep stage category:

Precision: It indicates the proportion of samples predicted to be positive that are actually positive, defined as:

$$\text{presicion} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

Recall: It indicates the proportion of samples that are actually positive and that are correctly classified as positive by the model. It is defined as:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

F1 Score: It is the harmonic average of precision and recall, which can balance the model performance between precision and recall, and is defined as:

$$\text{F1} = \frac{2 \cdot \text{Pr} \cdot \text{Re}}{\text{Pr} + \text{Re}} \tag{6}$$

Among them, TP (True Positive) is the number of samples correctly classified as this class, and FP (False Positive) is the number of samples misclassified as this class. FN (False Negative) is the number of samples that are actually positive but misclassified as other classes.

In addition, in order to evaluate the overall performance on all categories, this paper also uses the following four indicators:

Confusion Matrix: Provides detailed classification result statistics, showing the correct classification and misclassification between categories.

Accuracy: Indicates the overall correct classification ratio of the model for all samples, calculated as:

$$\text{ACC} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{TN}+\text{FP}} \tag{7}$$

Macro F1 (MF1): It represents the average of the F1 scores of each category, which is used to measure the balanced performance of the model on all categories. It is defined as follows:

$$\text{MF1} = \frac{1}{\text{n}} \sum_{\text{i}=1}^{\text{n}} \frac{2 \cdot \text{Pr}_\text{i} \cdot \text{Re}_\text{i}}{\text{Pr}_\text{i} + \text{Re}_\text{i}} \tag{8}$$

Kappa Coefficient: It is used to measure the difference between the accuracy of the classification model and random classification. The higher the Kappa value, the better the classification effect.

$$\text{Kappa} = \frac{P_0 - P_e}{1 - P_e} \tag{9}$$

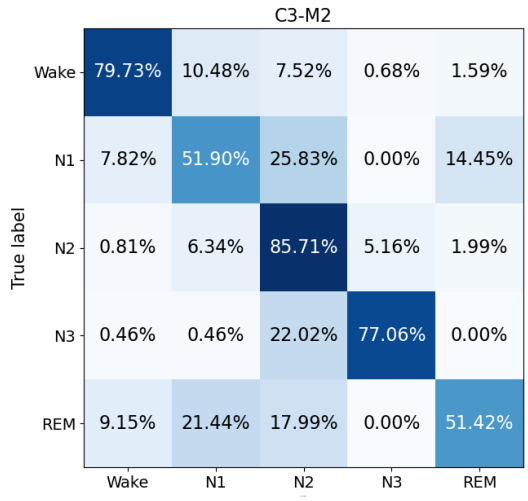
Among them, TN (True Negative) is the number of samples correctly classified as negative. P_0 represents the actual observed accuracy, that is, the classification accuracy of the model. P_e represents the expected accuracy, that is, the prediction accuracy of the model under purely random conditions.

E. Result and Discussion

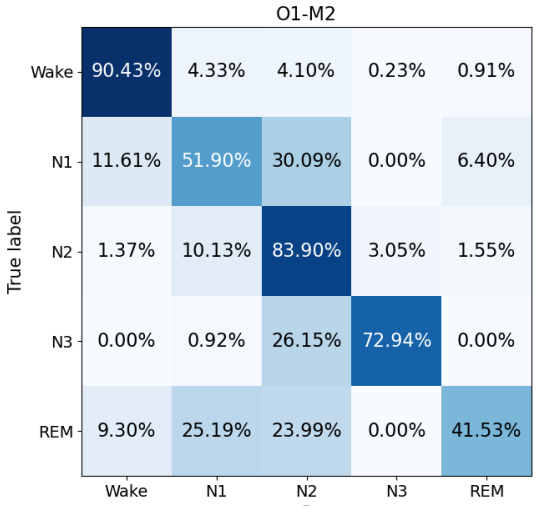
We initially evaluated the model using four unilateral EEG channels (O1-M2, C3-M2, F3-M2, E1-M2) without data augmentation. As shown in Fig. 4, the model achieved high accuracy in classifying Wake and N2 stages. However, performance on N1, N3, and REM stages was comparatively weaker due to their less distinct EEG features and the imbalanced class distribution. These results reflect the importance of spatial information in EEG-based sleep staging. For example, frontal channels (e.g., F3-M2) are more sensitive to transitions into light sleep, while occipital channels (e.g., O1-M2), despite reduced visual input during sleep, may contribute to REM detection due to their role in dream-related activity.

To improve performance, we incorporated all eight EEG channels and applied Borderline-SMOTE to address class imbalance, particularly for N3 and REM stages. As shown in Fig. 5, the multi-channel augmented model demonstrated significant improvements across all sleep stages, especially for N1 and REM. This enhancement can be attributed to three key factors: the inclusion of more spatially diverse EEG data, allowing the model to capture a broader range of sleep-related brain activity, the cross-layer connection module, which preserved low-level features critical for fine-grained classification; and the targeted augmentation strategy, which increased the representation of difficult boundary samples and improved the model's ability to generalize across sleep stage transitions.

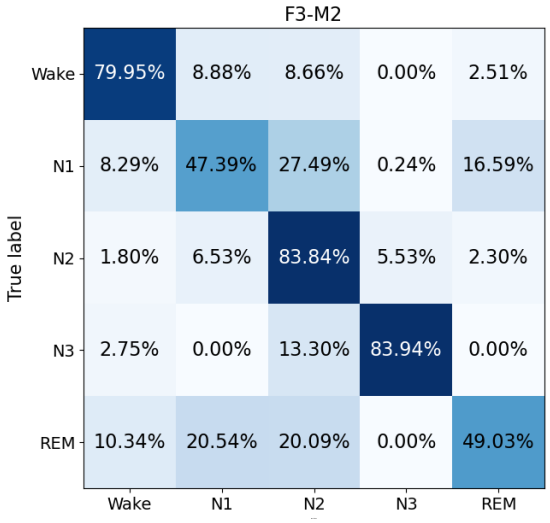
These results demonstrate that our proposed CrossLUSleepNet not only achieves high accuracy in standard scenarios but also maintains robustness and adaptability in challenging, imbalanced clinical data environments.



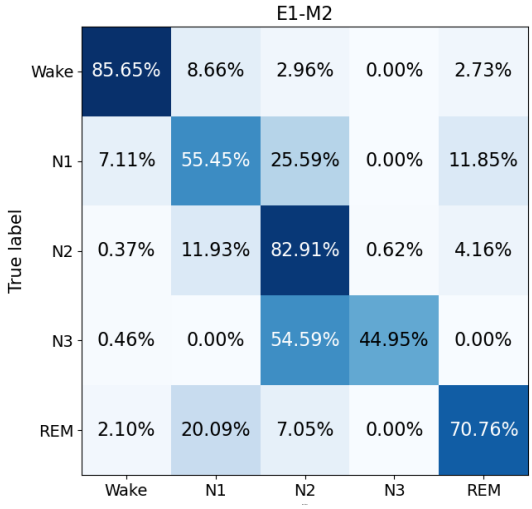
(a)



(b)



(c)



(d)

Fig. 4. Confusion matrix of unilateral EEG channels O1-M2, C3-M2, F3-M2, and E1-M2 at different sleep stages under the model test in this paper, respectively a, b, c, and d. Each cell in the matrix represents the predicted percentage of each actual category, represented by the color depth in the matrix. The rows correspond to different sleep stages, including W, N1, N2, N3, and REM, while the columns represent the stages predicted by the model. The darker the color, the higher the consistency between the predicted results and the actual categories, and the diagonal indicates the accuracy of classification at each stage. The confusion matrix helps identify the advantages and disadvantages of the model at different stages.

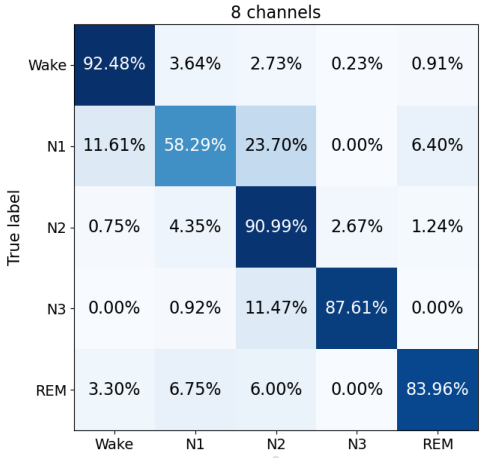


Fig. 5. Confusion matrix of different sleep stages based on 8-channel signal model testing.

We compared the proposed Cross-LUSleepNet with two representative sleep staging models DeepSleepNet and TransUSleepNet (which also employs a U-Net structure) on both clinical sleep patient datasets and public datasets. The results, presented in Table 2, illustrate the classification performance of each model across different datasets.

To validate the effectiveness of each module in the proposed model, we designed two benchmark frameworks for comparison. Framework 1 employs only the U-Net architecture for sleep stage classification, while Framework 2 integrates U-Net with a Bi-LSTM layer. To assess the impact of different model components on classification performance, we calculated the relative errors of key evaluation metrics between each framework and the full model. As shown in Table 3, Relative Error 1 denotes the discrepancy between Framework 1 and the proposed model, while Relative Error 2 reflects the difference between Framework 2 and the

proposed model.

As evidenced by Table 2, Cross-LUSleepNet achieves superior performance on clinical datasets, outperforming previous methods. Many existing models rely on public datasets consisting of healthy adults or patients with mild insomnia, which often fail to represent the complexity of real-world clinical PSG data. In contrast, clinical recordings are prone to issues such as electrode detachment, motion artifacts, unstable baselines, and high electrode impedance, which can result in partial data loss and increased variability. Additionally, clinical sleep data is often affected by factors like sleep fragmentation and arousals, further complicating model training and evaluation.

Unlike prior approaches, the proposed model directly learns from raw EEG and EOG signals, capturing multi-scale sleep transition patterns and enhancing feature representation through the combination of U-Net, Bi-LSTM, and a cross-layer connection module. This design enables the model to extract critical temporal and spatial features for more accurate sleep stage classification. As a result, Cross-LUSleepNet demonstrates notable improvements, particularly in challenging stages such as N1 and REM. The N1 stage, which often has the fewest samples and is prone to misclassification due to similarity with REM, showed a significant increase in F1 score. Overall, the model achieved a staging accuracy of 82.5% on raw clinical data, highlighting its strong generalization ability and robustness under real-world conditions.

From Table 3, it is clear that the proposed model, which integrates Bi-LSTM modules and cross-layer connections, achieves higher accuracy and macro F1 scores compared to the benchmarks. Across the different sleep stages, all three models exhibit similar performance in classifying stage W, likely due to the distinct and dominant waveform features of this stage. However, more notable differences emerge in stage N1 classification: Framework 2 and the proposed model demonstrate comparable performance with F1 scores exceeding 55%, reflecting the benefit of Bi-LSTM modules in enhancing the global contextual representation of sleep stages. In contrast, Framework 1 attains a lower F1 score of 52.7% for stage N1, indicating less effective modeling of temporal dependencies.

Overall, the results suggest that the combined use of Bi-LSTM and cross-layer connection modules significantly improves sleep staging accuracy compared to employing either module alone or using a simpler network design. As shown in Table 2, the sleep staging model developed in this study achieved an overall F1 score of 78.6%, representing a 2% to 5% improvement, and an accuracy of 84.7%, increasing by 2% to 7%, relative to comparable models. The proposed model consistently outperforms existing methods of similar architecture, particularly excelling in challenging stages such as N1, N3, and REM. Furthermore, its F1 score during training surpasses those of baseline approaches, confirming its superior classification capability.

TABLE II
PERFORMANCE OF DIFFERENT NETWORKS ON DIFFERENT DATASETS

Model	Sleep EDF			Clinical Data		
	ACC	MF1	Kappa	ACC	MF1	Kappa
Cross-LUSleepNet	87.65%	79.41%	0.79	**82.66%**	78.63%	0.79
DeepSleepNet	82.01%	76.05%	0.76	76.00%	70.90%	0.72
TransUSleepNet	**89.83%**	81.56%	0.85	78.54%	75.94%	0.77

TABLE III
RELATIVE ERROR OF INDICATORS BETWEEN FRAMEWORK 1, FRAMEWORK 2 AND THE PRESET MODEL

	Frame-work 1	**Frame-work 2**	**Proposed Model**	**Relative Error 1**	**Relative Error 2**
F1-W/%	86.2	88.2	89.4	3.2	1.2
F1-N1/%	52.7	57.8	60.1	7.4	2.3
F1-N2/%	82.5	84.2	85.8	3.3	1.6
F1-N3/%	79.2	84.6	89.8	10.6	5.2
F1-REM/%	80.5	84.2	87.6	7.1	3.4
MF1/%	73.8	76.4	78.6	4.8	2.2
ACC/%	76.2	79.8	82.6	6.4	2.8

V. Conclusions

This paper presents a novel deep learning approach, Cross-LUSleepNet, which integrates a U-shaped network architecture with bidirectional long short-term memory (Bi-LSTM) modules to effectively learn long-range temporal features from EEG signals. The model incorporates a specially designed cross-layer connection module that enhances feature continuity extraction by integrating multi-level features within the U-shaped structure. This module enables efficient information flow between adjacent layers and fuses features across different scales, thereby improving the global feature representation while retaining detailed local information. The embedded Bi-LSTM modules further capture long-term dependencies in temporal sequences, enhancing the model's ability to represent contextual information and improving its understanding of stage transitions. By mitigating feature loss and reinforcing temporal continuity, the model is better equipped to recognize the dynamic patterns inherent in sleep stage progression. Compared to existing state-of-the-art methods employing U-shaped architectures in sleep staging tasks, Cross-LUSleepNet demonstrates superior performance in both quantitative metrics and qualitative evaluation. Additionally, the model exhibits strong generalization capability, making it well-suited for real-world clinical applications involving complex and variable EEG data.

References

[1] D. Berg *et al.*, "Prodromal Parkinson disease subtypes—key to understanding heterogeneity," *Nature Reviews Neurology*, vol. 17, no. 6, pp. 349-361, 2021.

[2] P. Chriskos, C. A. Frantzidis, C. M. Nday, P. T. Gkivogkli, P. D. Bamidis, and C. Kourtidou-Papadeli, "A review on current trends in automatic sleep staging through bio-signal recordings and future challenges," *Sleep medicine reviews*, vol. 55, p. 101377, 2021.

[3] R. B. Berry *et al.*, "AASM scoring manual updates for 2017 (version 2.4)," vol. 13, ed: American Academy of Sleep Medicine, 2017, pp. 665-666.

[4] J. Liu *et al.*, "Automatic sleep staging algorithm based on random forest and hidden Markov model," *Computer Modeling in Engineering & Sciences*, vol. 123, no. 1, pp. 401-426, 2020.

[5] H. Ghimatgar, K. Kazemi, M. S. Helfroush, and A. Aarabi, "An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov Model," *Journal of neuroscience methods*, vol. 324, p. 108320, 2019.

[6] B. Yang, X. Zhu, Y. Liu, and H. Liu, "A single-channel EEG based automatic sleep stage classification method leveraging deep one-dimensional convolutional neural network and hidden Markov model," *Biomedical Signal Processing and Control*, vol. 68, p. 102581, 2021.

[7] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Computing and Applications*, vol. 28, pp. 2959-2978, 2017.

[8] I. N. Yulita, R. Rosadi, S. Purwani, and M. Suryani, "Multi-layer perceptron for sleep stage classification," in *Journal of Physics: Conference Series*, 2018, vol. 1028, no. 1: IOP Publishing, p. 012212.

[9] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Computers in biology and medicine*, vol. 106, pp. 71-81, 2019.

[10] J. Zhang, R. Yao, W. Ge, and J. Gao, "Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG," *Computer methods and programs in biomedicine*, vol. 183, p. 105089, 2020.

[11] H. Phan *et al.*, "L-SeqSleepNet: Whole-cycle long sequence modeling for automatic sleep staging," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 10, pp. 4748-4757, 2023.

[12] H. Phan, K. Mikkelsen, O. Y. Chén, P. Koch, A. Mertins, and M. De Vos, "Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 8, pp. 2456-2467, 2022.

[13] J. Sun *et al.*, "START: Automatic sleep staging with attention-based cross-modal learning transformer," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2023: IEEE, pp. 1421-1428.

[14] Y. Guo, M. Nowakowski, and W. Dai, "FlexSleepTransformer: a transformer-based sleep staging model with flexible input channel configurations," *Scientific Reports*, vol. 14, no. 1, p. 26312, 2024.

[15] Y. Dai *et al.*, "Multichannelsleepnet: A transformer-based model for automatic sleep stage classification with psg," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 9, pp. 4204-4215, 2023.

[16] J. Li, B. Hu, Z.-H. Guan, and B. Zhang, "Sleep Stage Classification with Spiking Neural Networks and Transformers using Multi-Channel EEG Data," in *2024 43rd Chinese Control Conference (CCC)*, 2024: IEEE, pp. 8345-8350.

[17] Z. Chen *et al.*, "Automated sleep staging via parallel frequency-cut attention," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1974-1985, 2023.

[18] R. Soleimani *et al.*, "Advances in modeling and interpretability of deep neural sleep staging: A systematic review," *Physiologia*, vol. 4, no. 1, pp. 1-42, 2023.

[19] R. N. Sekkal, F. Bereksi-Reguig, D. Ruiz-Fernandez, N. Dib, and S. Sekkal, "Automatic sleep stage classification: From classical machine learning methods to deep learning," *Biomedical Signal Processing and Control*, vol. 77, p. 103751, 2022.

[20] Q. Gao and K. Wu, "Automatic sleep staging based on power spectral density and random forest," *Sheng wu yi xue Gong Cheng xue za zhi= Journal of Biomedical Engineering= Shengwu Yixue Gongchengxue Zazhi*, vol. 40, no. 2, pp. 280-285, 2023.

[21] X. Yu, C. Liu, J. Dai, J. Li, J. Wang, and F. Hou, "Epilepsy electroencephalogram signal analysis based on improved K-nearest neighbor network," *Sheng wu yi xue Gong Cheng xue za zhi= Journal of Biomedical Engineering= Shengwu Yixue Gongchengxue Zazhi*, vol. 33, no. 6, pp. 1039-1045, 2016.

[22] S. Mousavi, F. Afghah, and U. R. Acharya, "SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PloS one*, vol. 14, no. 5, p. e0216456, 2019.

[23] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 25, no. 11, pp. 1998-2008, 2017.

[24] A. Supratak and Y. 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)*, 2020: IEEE, pp. 641-644.

[25] H. Phan, O. Y. Chén, M. C. Tran, P. Koch, A. Mertins, and M. De Vos, "XSleepNet: Multi-view sequential model for automatic sleep staging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5903-5915, 2021.

[26] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," *Advances in neural information processing systems*, vol. 32, 2019.

[27] Z. Jia *et al.*, "GraphSleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification," in *Ijcai*, 2020, vol. 2021, pp. 1324-1330.

[28] D. Zhang, Y. She, J. Sun, X. Yang, X. Zeng, and W. Qin, "SwinSleep: A Deep Learning Framework Advancing Overnight Sleep Staging towards Clinical Practice," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[29] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE access*, vol. 9, pp. 82031-82057, 2021.

[30] J. You, D. Jiang, Y. Ma, and Y. Wang, "SpindleU-Net: An adaptive u-net framework for sleep spindle detection in single-channel EEG," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 29, pp. 1614-1623, 2021.

[31] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on U-net: A review," *Journal of Imaging Science & Technology*, vol. 64, no. 2, 2020.

[32] T. I. Toma and S. Choi, "An end-to-end multi-channel convolutional Bi-LSTM network for automatic sleep stage detection," *Sensors*, vol. 23, no. 10, p. 4950, 2023.

[33] M. Längkvist and A. Loutfi, "A deep learning approach with an attention mechanism for automatic sleep stage classification," *arXiv preprint arXiv:1805.05036*, 2018.

[34] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the

slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185-1194, 2000.