Applied Intelligence (2025) 55:741
https://doi.org/10.1007/s10489-025-06587-y

SCANSleepNet: A spatial-channel attention network for sleep stage classification

Yuyun Liu[1] · Qilei Li[2] · Mingliang Gao[1] · Xiangyu Guo[1] · Wenzhe Zhai[1]

Accepted: 20 April 2025

Abstract
Sleep stages refer to the distinct processes within a person's sleep cycle. They are essential for assessing mental and physical health. Existing sleep stage classification models typically improve performance through increased computation complexity or be trained with more labeled data. These models may result in overly heavy models that are unrealistically not applicable in real-world scenarios. To address this issue, this paper proposes a Spatial-Channel Attention Network for Sleep Stage Classification (SCANSleepNet). This framework is built on the time-frequency characteristics of EEG signals and the conversion rules of sleep stages. It constructs a lightweight deep-learning system that integrates multi-scale frequency analysis and dynamic feature enhancement. Its improvements lie in two main aspects. First, a Spatial-Channel Dimensional Attention (SCDA) block is designed to model the dynamic transition among stages while requiring fewer parameters. Second, a weighted cross-entropy loss function is introduced to address class imbalance without dependence on extra data augmentation. It enhances the model's lightness and suitability for clinical applications. Experimental results show that the SCANSleepNet achieved 85.52% accuracy in the Fpz-Cz channel and 82.16% in the Pz-Oz channel on the Sleep-EDF dataset. It makes a good balance between classification accuracy and efficiency.

Keywords Deep learning · Sleep stage classification · Attention mechanism · Multiscale atrous convolution

1 Introduction

Monitoring and evaluating sleep quality are essential for maintaining overall health and cognitive function [1–4]. Automatic sleep stage classification is vital to assess sleep quality and detect potential sleep disorders. Advances in deep learning technology have facilitated the development of automated sleep stage classification algorithms that outperform conventional machine learning approaches and even human specialists in accuracy and efficiency [5, 6]. Sleep experts typically use multiple channels polysomnography, *e.g.,* electroencephalogram (EEG), electrooculogram, electromyogram, and electrocardiogram, to determine the stages of sleep [7]. Among these channels, single-channel EEG has gained attention in sleep assessment due to its flexibility and convenience.

The single-channel EEG recordings are typically segmented into 30-second intervals, with each segment manually reviewed by sleep experts to classify them into six stages, namely Wakefulness (W), Rapid Eye Movement (REM), and four non-REM stages (N1, N2, N3, and N4) [8]. Significant differences in the spectral characteristics of brain electrical signals are presented in each stage of the human sleep cycle. In the initial N1 stage of sleep, the EEG is mainly characterized by theta waves (4-8Hz) accompanied by a small amount of alpha waves (8-12Hz). After entering the N2 stage, the amplitude of the EEG signal is significantly enhanced. Not only do the theta wave characteristics continue to intensify, but K-complex waves also appear. The deep sleep N3 stage is dominated by a mixture of theta and delta waves (0-4Hz), with a significant increase in the proportion of delta waves. By the N4 stage, the frequency of the EEG signal is further reduced to the range of 0.5-2Hz. The REM sleep period includes activity in the sigma waves (12-15Hz), beta waves (15-30Hz), and gamma waves (greater

Yuyun Liu and Qilei Li contributed equally to this work.

✉ Mingliang Gao
mlgao@sdut.edu.cn

[1] School of Electrical and Electronic Engineering, Shandong University of Technology, 255000 Zibo, China

[2] School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, UK

Published online: 18 June 2025

than 30Hz). The Wakefulness state primarily exhibits beta-wave activity. This manual procedure is meticulous, dull, and time-consuming. Therefore, an automated system for classifying sleep stages is essential to support sleep experts.

In recent years, the rapid development of deep learning technology has injected new vitality into the research of sleep stage classification. However, current models have many limitations when they process complex sleep signals and practical application scenarios. Previous research divides end-to-end models into two main categories: methods based on recurrent neural networks (RNNs) and their variants, and composite models combined with attention mechanisms. Although RNN-based methods (such as LSTM and Bi-LSTM) have demonstrated excellent performance in sleep stage classification [5, 6, 9, 10]. The inherent problems with long sequences, such as gradient decay or explosion, restrict the stability and performance of the models. In addition, the serial computation results in low efficiency during model optimization and makes it challenging to meet the requirements of practical applications. To improve the computational efficiency of the models, Zhao et al. [11] proposed a multi-task deep learning model integrating sequence signal reconstruction. However, it has limitations in feature extraction as they don't fully account for model sensitivity to different frequency components.

To solve the data class imbalance issue, Mousavi et al. [12] proposed a Synthetic Minority Over-sampling Technique (SMOTE). Chen et al. [13] optimized signal generation quality with the multi-scale convolution and deep spectrum interaction modules, and proposed an enhancement method via generated data label reconstruction and time-domain reorganization. However, the model complexity limits practical deployment. Ying et al. [14] built a single-channel dual-stream network (Ds-ASSNet). It extracts fine and coarse-grained temporal features with a multi-scale 1D-CNN. The features are combined with the Bidirectional Gated Recurrent Unit (BiGRU) to capture temporal dependencies. With the rapid advancement of artificial intelligence technology, explainable artificial intelligence has become a focal point in both academic and industrial research [15]. The lack of model interpretability remains a major barrier to the wide application of transfer learning.

To address the issues of data imbalance, insufficient exploitation of the frequency sensitivity of EEG signals, limited model interpretability, and high computational resource demands, we propose a lightweight deep learning framework for sleep stage classification, termed spatial-channel attention Network (SCANSleepNet). This framework captures time-domain EEG signal features at multiple scales through dilated convolution operations. It performs fine-grained feature encoding and extraction of various frequency components, thereby fully leveraging the frequency sensitivity of EEG signals. At the same time, a Spatial-Channel Dimensional Attention (SCDA) block module is introduced to model the dynamic transition relationships between sleep stages and capture important features and potential changes. Through optimization of the network structure and attention mechanism, SCANSleepNet reduces the demand for computing resources and significantly improves operational efficiency. The proposed model provides an efficient, lightweight, and easy-to-deploy solution for the sleep stage classification task. Comprehensive experiments on the Sleep-EDF dataset showcase SCANSleepNet's superior performance, especially on the Fpz-Cz and Pz-Oz channels. This demonstrates its effectiveness in real-world sleep stage classification tasks. The key contributions of this work are outlined as follows.

- We propose a SCANSleepNet for sleep stage classification using raw single-channel EEG signals. It can effectively learn and integrate spatial and channel information with varying frequency components through CNNs and attention modules.
- We introduce an SCDA block to capture temporal correlations in both the feature space and channel dimensions of complex time series.
- We developed a loss function termed weighted focal cross-entropy loss to effectively address class imbalance without dependence on data augmentation. It considers both the differences in class numbers and the learning difficulties across categories.

2 Related work

2.1 Convolutional neural networks

Identifying sleep stages is of significant importance for diagnosing sleep disorders and assessing psychological conditions. In early studies, methods were employed to classify sleep stages through Support Vector Machines (SVM) and Random Forests (RF). However, these approaches require extensive prior knowledge of feature engineering. Therefore, researchers have increasingly adopted deep learning networks for sleep stage classification. Tsinalis et al. [16] utilized Convolutional Neural Networks to learn deep features for classification from single-channel EEG signals without prior knowledge. Chriskos et al. [17] utilized the SMOTE algorithm to enable CNNs to appropriately consider small samples and make accurate identifications in the medical domain. Sokolovsky et al. [18] employed deeper network architectures to classify sleep stages with multi-channel signals.

The CNN models excel in sleep stage classification and show their ability to effectively recognize features within EEG signals without the need for manual feature extraction.

However, sleep stage signals are inherently continuous over time and exhibit temporal dependencies. Researchers have proposed RNNs to address this challenge. RNNs excel at learning sequential patterns within data. They are particularly suitable for tasks that require the comprehension of temporal contexts, such as the analysis of continuous signals in sleep stage classification.

2.2 Recurrent neural networks

Recurrent Neural Networks (RNNs) are particularly adept at handling time series due to their distinctive structure and design advantages. Yin et al. [19] utilized wearable medical sensors to collect biological signal data and applied LSTM for analysis. DeepSleepNet [6] is an advanced method that employs deep learning techniques for sleep stage classification. This approach applies convolutional neural networks to derive time-invariant characteristics from single-channel electroencephalogram data. Meanwhile, it uses bidirectional long-term and short-term memory networks (Bi-LSTM) to capture the transition patterns that exist between various stages of sleep. In CCRRSleepNet [9], a hybrid architecture of CNNs and RNNs was built to capture high-order time-varying and time-invariant signal characteristics and model the relationships between sleep stages' transitions.

In sleep stage classification, RNNs are typically employed after CNNs to model the subtle transition relationships between different sleep stages. However, the "black box" characteristic of RNNs makes them insufficiently interpretable in the medical field. In addition, when dealing with long sequences, the model has inherent problems such as vanishing gradients or exploding gradients. These issues severely limit the stability and performance of the model. Moreover, the feature of serial computation reduces the efficiency of the training.

2.3 Attention mechanism

The rise of the Transformer architecture and its extensive applications in various fields [20, 21] have made it a research hotspot. Its remarkable sequence modeling capabilities and global attention mechanism have contributed to its popularity in multiple fields such as natural language processing and computer vision. Compared with traditional models, the Transformer architecture usually requires more parameters and computational resources to maintain its performance advantages. This makes the model increasingly burdensome and the trend stands in sharp contrast to the urgent need for lightweight models in various industries in recent years [22–24]. In practical applications, especially in scenarios with limited resources, the lightweight of models has become crucial.

Eldele et al. [25] proposed an AttnSleep model based on multi-resolution CNNs and attention mechanisms. The model first extracts low-frequency and high-frequency features from EEG signals and then captures temporal dependencies between features through a Time Context Encoder (TCE) equipped with multi-head attention mechanisms. Mousavi et al. [10] proposed a SleepEEGNet model that constructed a sequence-to-sequence model based on an encoder-decoder architecture. This approach first extracts invariant features from EEG signals, then processes these features through a bidirectional recurrent convolutional network encoder, and combines an attention module to enhance the representation of key features. These methods have limitations in feature extraction. They only extract specific features from raw time-domain signals and do not fully consider the model's sensitivity to different frequency components. Meanwhile, previous studies confirm the significance of spatio-temporal analysis for video data. Temporal analysis captures the temporal relations and changes in actions and locates key time points. Spatial analysis focuses on framing visual content to identify key regions and objects. Their combination helps understand the model's spatio-temporal focus and improves interpretability and performance [26]. To address this issue, we propose a method based on dilated convolution to expand the model's receptive field and capture broader spatial and temporal information. We extend spatio-temporal analysis to EEG signal processing and introduce a Spatial-Channel Dimensional Attention block. The spatial dimension captures the signal's spatial distribution via dilated convolutions. The channel dimension reflects its activity in different frequency bands. These two dimensions comprehensively capture key features and patterns for subsequent analysis and classification.

3 The proposed method

3.1 Overview of SCANSleepNet

The architecture of SCANSleepNet is shown in Fig. 1. It comprises three key modules, namely the Intra-Frame Feature Extraction (IFFE) module, Temporal Feature Capture (TFC) module, and classification module. The IFFE module extracts meaningful intra-frame features from the input data, which reflect the frequency information that remains constant over time. The (TFC) module encodes the extracted feature maps to capture and represent changes in the data across the time dimension. This allows the system to recognize and leverage temporal correlations. Finally, the classification module uses these extracted and encoded features to decode temporal information and perform precise classification tasks. This enables the accurate determination and analysis of the stages of sleep. Overall, this multi-layered processing architecture

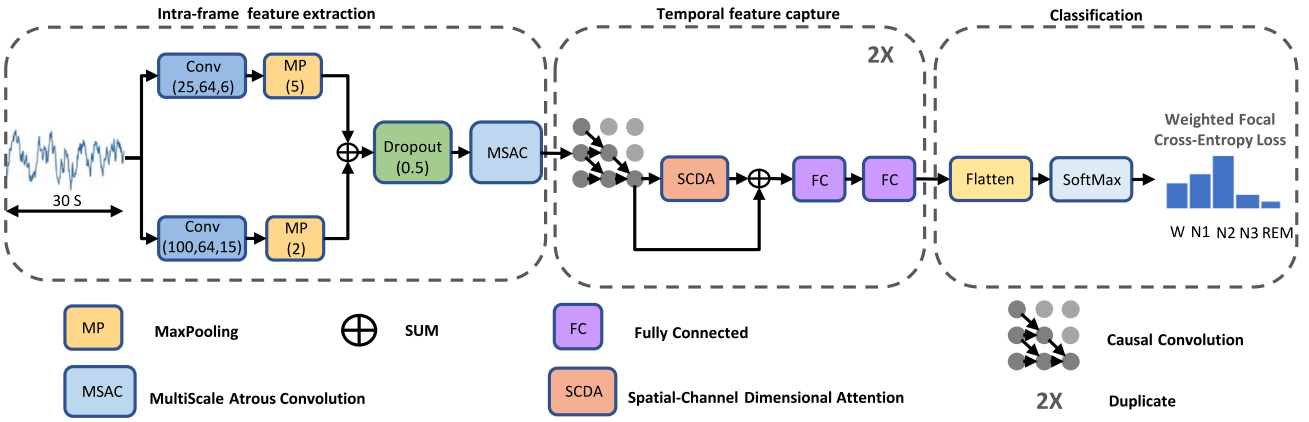


Fig. 1 Overall architecture of the SCANSleepNet

ensures the system's efficiency and accuracy in real-time monitoring and analysis of sleep patterns.

3.2 Intra-frame feature extraction module

This approach is designed to analyze non-stationary time-varying EEG signals and extract intra-frame relevant features from the input data. It employs a frame-level CNN architecture from CCRRSleep [9]. The use of convolutional kernels of different sizes helps to better extract frequency features of different sleep stages.

The Conv (25, 64, 1) denotes a convolutional layer with the kernel of size of 25 × 25, 64 filters, and a stride of 1. Likewise, MaxPooling (5) indicates a maximum pooling layer uses a kernel size of 5. Each convolutional layer includes a batch normalization layer [27] and utilizes Mish [28] as its activation function. It is defined as follows:

$$\text{Mish}(\mathbf{x}) = \mathbf{x} \cdot \tanh\left(\log\left(\mathbf{1} + \mathbf{e}^{\mathbf{x}}\right)\right). \tag{1}$$

In light of the aforementioned spectral characteristics, we select two distinct convolutional kernel sizes, namely 25 and 100, to construct a highly efficient feature extraction system. The rationale behind this selection is based on the inherent physical relationship among the sampling rate (F_s), the convolutional kernel size (K), and the target frequency (f). It can be formulated as:

$$f = \frac{F_s}{K}. \tag{2}$$

Under a 100 Hz sampling rate, when $K = 25$, the theoretically captured frequency is 4 Hz. It covers the theta wave band (4-8Hz) that spans N1-N3 stages. When $K = 100$, the theoretical frequency is 1Hz, which matches the low-frequency characteristics of the N3 stage dominated by delta waves (0-4Hz) and the N4 stage (0.5-2Hz). This dual-scale design enables targeted capture of critical frequency-domain feature differences in the non-REM cycle (N1-N4). Considering that the REM stage primarily involves high-frequency components above 12Hz (sigma/beta/gamma waves), this study did not adopt a specially designed convolutional kernel architecture to capture these signal features. There are two main advantages to the proposed module. On the one hand, through the convolutional kernel design with physical constraints, it is ensured that the feature extraction process has a clear physiological significance orientation. On the other hand, the size combination of 25/100 not only ensures the accurate capture of theta/delta waves but also takes into account the feature coverage efficiency of different sleep stages.

The features extracted from the two branches of the IFFE module are concatenated along the channel dimension. If $F_1 \in \mathbb{R}^{H\times W\times C_1}$ and $F_2 \in \mathbb{R}^{H\times W\times C_2}$ are the feature maps from the two branches, the concatenated feature map F_{concat} is given by:

$$F_{concat} = [F_1; F_2] \in \mathbb{R}^{H\times W\times (C_1+C_2)}. \tag{3}$$

Across both branches of the network, the pooling kernel size and the convolution stride before it remains consistent. The features that are extracted from the CNN branches are concatenated, and a dropout layer is applied with a retention rate of 0.5 to mitigate overfitting. Furthermore, the network incorporates MSAC block [9] to extract features. The structure of the MSAC block is shown in Fig. 2. The MSAC brock [9] employs convolutional layers with different depths and field sizes. These layers enable the comprehensive extraction of frequency components and features with complexities from sleep signals. Let the input feature map be X_{in} and the output feature map be X_{out}. In the diagram, the convolution operation is denoted as $Conv$, the concatenation operation as $Concat$, and the batch normalization and activation operation as Bn. The specific calculation steps are as

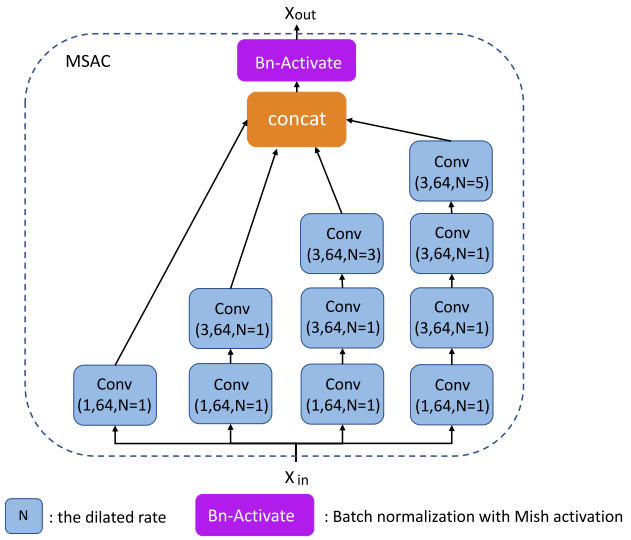


Fig. 2 Framework of the MSAC block

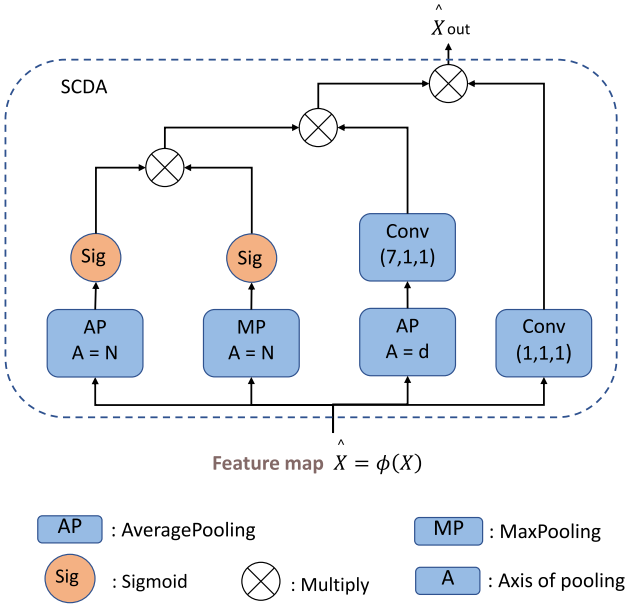


Fig. 3 Framework of the SCDA block

follows:

$$
\begin{aligned}
T_1 &= \mathrm{Conv}(1, 64, N = 1)(X_{\mathrm{in}}),\\
T_2 &= \mathrm{Conv}(3, 64, N = 1)(T_1),\\
T_3 &= \mathrm{Conv}(3, 64, N = 3)(T_2),\\
T_4 &= \mathrm{Conv}(3, 64, N = 5)(\mathrm{Conv}(3, 64, N = 1)(T_2)),\\
T_{\mathrm{c}} &= \mathrm{Concat}(T_1, T_2, T_3, T_4),\\
X_{\mathrm{out}} &= \mathrm{Bn}(T_{\mathrm{c}}),
\end{aligned}
\tag{4}
$$

where $Conv(a, b, N = c)$ represents a convolutional layer with a kernel size of $a \times a$, an output channel number of b, and a dilation rate of c. The final output feature map is generated by performing batch normalization on the concatenated feature map T_{concat} and using the Mish activation function.

3.3 Temporal feature capture module

The TFC module is designed to capture the temporal correlations within the input features. The TFC module comprises a causal convolution, a Spatial-Channel Dimensional Attention (SCDA) block, as well as a fully connected (FC) layer. The SCDA block is illustrated in Fig. 3. Additionally, two identical structures of the TFC module are stacked to generate the final features. The input to the TFC module is denoted as $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \in \mathbb{R}^{N \times d}$ and it consists of N features. Each feature has a length of d. Firstly, we apply causal convolutions to generate $\hat{\mathbf{X}}$ from $\mathbf{X}$, denoted as $\hat{\mathbf{X}} = \phi(\mathbf{X})$. Then $hat\ mathbf\ X$is inputted into the SCDA Block. SCDA block captures relevant features and attention weights from the input tensor. These operations help extract meaningful information and capture both channel-wise and spatial-wise correlations within the data. On the dimension N of the input feature map, average pooling and maximum pooling are applied to generate the compressed feature vectors $\hat{\mathbf{x}}_{\mathrm{avg\_N}} \in \mathbb{R}^{1 \times d}$ and $\hat{\mathbf{x}}_{\mathrm{max\_N}} \in \mathbb{R}^{1 \times d}$. Meanwhile, average pooling is performed on the feature dimension d to generate $\hat{\mathbf{x}}_{\mathrm{avg\_d}} \in \mathbb{R}^{N \times 1}$. A convolution operation is used to generate a high-order representation $\hat{\mathbf{x}}_{conv}$ with local correlation characteristics.

Subsequently, $\hat{\mathbf{x}}_{\mathrm{avg\_N}}$ and $\hat{\mathbf{x}}_{\mathrm{max\_N}}$ pass through two linear layers to produce $\hat{\mathbf{x}}_{\mathrm{avg\_fc}}$ and $\hat{\mathbf{x}}_{\mathrm{max\_fc}}$, respectively. Then, $\hat{\mathbf{x}}_{\mathrm{avg\_fc}}$ and $\hat{\mathbf{x}}_{\mathrm{max\_fc}}$ are multiplied together. The sigmoid function activates the linear layers.

$$
\begin{aligned}
\hat{\mathbf{x}}_{\mathrm{avg\_fc}} &= \mathrm{sigmoid}\left(\hat{\mathbf{x}}_{\mathrm{avg\_N}}\right),\\
\hat{\mathbf{x}}_{\mathrm{max\_fc}} &= \mathrm{sigmoid}\left(\hat{\mathbf{x}}_{\mathrm{max\_N}}\right),
\end{aligned}
\tag{5}
$$

where $\hat{\mathbf{x}}_{\mathrm{avg\_d}}$ is obtained with a convolutional layer of kernel size 7 and stride 1. The output is multiplied by $\hat{\mathbf{x}}_{\mathrm{avg\_fc}}$ and $\hat{\mathbf{x}}_{\mathrm{max\_fc}}$ to produce $\hat{\mathbf{x}}_{\mathrm{weight}}$. The $\hat{\mathbf{x}}_{\mathrm{weight}}$ are multiplied with $\hat{\mathbf{x}}_{\mathrm{conv}}$ to produce the output of the SCDA block. The results of the SCDA block are added to those of the causal convolution. The combined output of the TFC module is then passed through a fully connected layer to produce the final result.

3.4 Weighted focal cross-entropy loss

There are two main challenges in classifying sleep stage samples. First, distinguishing between different class samples is challenging. For instance, the N1 stage is often misclassified due to its transitional nature and similarity to adjacent stages, with its short duration and limited sample size further complicating classification. Second, there is variability among samples within the same class, primarily influenced by the

importance of intra-class features and noise. Individual differences, measurement errors, and environmental factors lead to diverse sample characteristics. The model must exhibit higher robustness to handle such complex data.

Prior studies have addressed the data imbalance issue through the focal loss function [9]. The mechanism of focal loss adjusts sample weights to prioritize challenging samples. This enhances training performance under class imbalance. The focal loss introduces weights for difficult-to-easy samples and dynamically adjusts the weights based on the model's predictions. Misclassified samples receive higher weights, which enables the model to focus more on these challenging instances during subsequent training. However, since the focal loss function was initially created to reduce the emphasis on easily classified samples and prioritize challenging samples, it may not sufficiently penalize the misclassification of easily classified samples in certain cases. This can lead to suboptimal performance when the differences between classes are small or when there is noise interference.

To overcome this limitation, we propose a weighted focal cross-entropy loss function that integrates both the focal loss and the class-aware loss function. This allows the model to effectively handle class imbalance and focus on learning challenging samples during the optimization process. The computation formula for the Weighted focal cross-entropy loss function is as follows:

$$L\left(\mathbf{y},\hat{\mathbf{y}}\right) = -\left(1-\beta\right)\frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{K} w_k \mathbf{y}_{k_i}\log\left(\hat{\mathbf{y}}_{k_i}\right) - \beta\left(1-\hat{\mathbf{y}}\right)^{r}\log\left(\hat{\mathbf{y}}\right),$$
$$\text{S.T. } \omega_k = \mu_k \cdot \max\left(1, \log\left(M/M_k\right)\right), \tag{6}$$

where ω_k denotes the weight allocated to the class k. M denotes the number of samples in a specific sleep stage type, and M_k represents the total number of samples in all sleep stages.

The combination of focal loss with class-specific cross-entropy loss yields the weighted focal cross-entropy loss function. It effectively addresses the issue of class imbalance during the optimization process. The weighted focal cross-entropy loss function assigns higher weights to difficult-to-classify samples. This ensures that the model pays more attention to these challenging instances. Thus, it improves the model's ability to recognize minority-class samples. The weighted focal cross-entropy loss function also emphasizes difficult samples, which prevents the model from overfitting to the majority class. As a result, the model learns more representative features during the optimization phase, which improves performance in practical applications. Finally, the Adam optimizer [29] is adopted to update the model parameters by minimizing the Weighted focal cross-entropy loss function.

4 Experimental results and analysis

4.1 Evaluation metrics

To gain a comprehensive understanding of the model's performance, several evaluation metrics were employed. For per-class evaluation, metrics such as precision, recall, and F1-score were considered. Precision measures the accuracy of the model's predictions, recall focuses on the proportion of actual positive samples that were captured by the model, and the F1-score provides a harmonic mean of precision and recall, reflecting the model's overall classification capability. Additionally, to assess the model's effectiveness in overall classification tasks, we used Macro-averaging F1-score (MF1), overall Accuracy, and Cohen's Kappa coefficient (k). The MF1 represents the average of the F1-scores across different classes and offers insight into the model's balanced performance across all categories. Overall accuracy reflects the proportion of correctly predicted samples out of all samples. Cohen's Kappa coefficient measures the agreement between the model's predictions and the actual labels. It accounts for the effect of random guessing and offers a more reliable evaluation of classification performance. These results will help optimize the model to enhance its accuracy and robustness in sleep stage classification tasks.

4.2 Benchmark dataset

The Sleep-EDF [30] dataset is adopted for performance assessment. This dataset features a sleep-cassette subset comprising 39 records from 20 subjects. The records include overnight polysomnographic sleep data from healthy Caucasian individuals aged 25 to 101 without sleep-related medications. Based on the R&K standard, expert annotations provide a reliable classification of the different sleep stages. In this study, stages S3 and S4 were combined into stage N3, following American Academy of Sleep Medicine (AASM) guidelines. The EEG data analyzed were from the Fpz-Cz and Pz-Oz channels, sampled at 100 Hz.

To avoid any overlap between training and testing sets for the same subjects and to ensure fair and unbiased comparisons with prior research, we followed the approach outlined in CCRRSleep [9]. Before and after sleep, we selected the Wake (W) period data from 30 minutes.

4.3 Experimental results

Table 1 presents the data analysis results for the Fpz-Cz channel and Pz-Oz channel. It includes a confusion matrix derived from 20-fold cross-validation. The matrix displays true and predicted class labels, while precision, recall, and F1 score values are provided for each category. In the Fpz-Cz channel, the Pre of the W stage reaches 90.03%, and the Pre of

Table 1 Performance of SCANSleepNet on Fpz-Cz and Pz-Oz channels

Stage	Fpz-Cz Predicted					Performance(%)			Pz-Oz Predicted					Performance(%)		
	W	N1	N2	N3	REM	Pre	Re	F1	W	N1	N2	N3	REM	Pre	Re	F1
W	**7420**	349	151	25	208	90.03	91.01	90.52	**7016**	286	104	13	288	85.18	91.03	88.01
N1	463	**926**	582	4	829	50.03	39.79	44.39	714	**512**	566	4	1007	42.77	18.27	25.60
N2	182	239	**15996**	529	853	89.34	89.61	89.47	229	192	**15414**	809	1154	87.29	86.61	86.95
N3	25	0	471	**5204**	3	90.28	90.76	90.52	29	2	804	**4853**	15	85.40	85.10	85.25
REM	152	337	704	2	**6522**	77.50	80.86	79.15	249	205	770	4	**6489**	72.48	84.09	77.85

Bold numbers represent correct classifications

the N3 stage is 90.28%, demonstrating outstanding classification performance; the N2 stage also has a Pre of 89.34%. In the Pz-Oz channel, the N2 stage shows a Pre of 87.29%. However, the N1 stage classification accuracy in both channels is relatively low: the Fpz-Cz N1 stage has a Pre of only 50.03%, while the Pz-Oz N1 stage Pre drops to 42.77%.

The comparison of SCANSleepNet with other methods using the Fpz-Cz EEG channel is listed in Table 2. One can see that SCANSleepNet shows excellent per-class F1-scores. In the wakefulness (W1) stage, it reaches an F1-score of 90.52%. This indicates its significant advantage in identifying wakefulness. In the N2 sleep stage, it achieves an F1-score of 89.61%. In the N3 sleep stage, it achieves an F1-score of 90.76%. The model can effectively capture the unique characteristics of these sleep stages. This reflects its adaptability and precision in dealing with different sleep stages. In terms of overall metrics, SCANSleepNet has an accuracy of 85.52%. It has an MF1 score of 78.31%. There is a good balance between them. It has a strong generalization ability across multiple sleep stages. The Kappa coefficient is 0.80. This further proves its robustness in sleep stage classification.

The comparison of SCANSleepNet with other methods using the Pz-Oz EEG channel is shown in Table 3. In the wakeful stage, it maintains an F1-score of 88.01%. This highlights its versatility for different electrode placements. SCANSleepNet stands out as a top-performing solution compared with other models. Its performance surpasses well-established models. These models include SleepEEG [10], ResnetLSTM [31], MultitaskCNN [32], and CCRRSleep [9]. The computational complexity and FLOPs compared with other models are shown in Table 4. It shows that SCANsleepNet improves computational efficiency while maintaining classification accuracy. It reduces FLOPs to 31.15 million (49.7% less than AttnSleep's 61.97 million) which demonstrates substantial resource savings for edge applications. On Fpz-Cz EEG channels, SCANsleepNet achieves 85.52% sleep staging accuracy (1.12% higher than AttnSleep) with a Kappa coefficient of 0.80. The balance of efficiency and performance makes it ideal for sleep stage classification tasks.

5 Ablation study

5.1 Component analysis

SCANSleepNet incorporates the IFFE module, TFC module, and WFCE loss function. To assess the effectiveness of each module within SCANSleepNet, we conducted an ablation study using the Sleep-EDF-20 dataset's Fpz-Cz channel. Specifically, we evaluated the TFC module and the influence of the weighted focal cross-entropy loss function. Counterparts are denoted as follows.

Table 2 Comparison of SCANSleepNet with other methods using Fpz-Cz channel

Method	Per-Class F1-score					Overall Metrics		
	W1	N1	N2	N3	REM	Accuracy	MF1	k
DeepSleep [6]	84.70	46.60	85.90	84.80	82.40	81.90	76.60	0.76
SleepEEG [10]	89.40	44.40	84.70	84.60	79.60	81.50	76.60	0.75
ResnetLSTM [31]	86.50	28.40	87.70	89.80	76.20	82.50	73.70	0.76
MultitaskCNN [32]	87.90	33.50	87.50	85.80	80.30	83.10	75.00	0.77
AttnSleep [25]	89.70	42.60	88.80	90.20	79.00	84.40	78.10	0.79
CCRRSleep [9]	89.01	51.73	87.25	88.20	82.86	84.29	79.81	0.78
U-Time [33]	87.00	52.00	86.00	84.00	84.00	–	79.00	–
SCANSleepNet	90.52	39.79	89.61	90.76	80.86	85.52	78.31	0.80

The best and second-best results are marked in red and blue, respectively

Table 3 Comparison of SCANSleepNet with other methods using Pz-Oz channel

Method	Per-Class F1-score					Overall Metrics		
	W1	N1	N2	N3	REM	Acc	MF1	k
DeepSleep [6]	88.10	37.00	82.70	77.30	80.30	79.80	73.10	0.72
ResnetLSTM [31]	85.60	24.90	88.90	79.20	86.30	81.00	73.60	–
CCRRSleep [9]	86.01	41.54	84.87	80.97	79.56	80.31	74.59	0.73
SCANSleepNet	88.01	25.60	86.95	85.25	77.85	82.16	72.73	0.75

The best and second-best results are marked in red and blue, respectively

- **IFFE only:** Using only the IFFE module resulted in an accuracy of 84.02%, an MF1 score of 76.73, and a k value of 0.78. This establishes a baseline for evaluating the impact of temporal feature capture.
- **IFFE + One TFC:** Adding a single layer of TFC to IFFE improved the accuracy to 84.07%, the MF1 score to 78.00, and the k value to 0.78. This highlights the significant role of temporal feature capture in enhancing model performance.
- **IFFE + Two TFC:** Utilizing two layers of TFC with the focal loss function achieved an accuracy of 84.60%. It also resulted in an MF1 score of 77.39, and a k value of 0.79. This indicates further performance improvements, showcasing the benefits of additional temporal feature capture layers.
- **IFFE + Two TFC with Class-aware loss:** When employing the class-aware (CA) loss function, the model achieved an accuracy of 84.21%, an MF1 score of 76.20, and a k value of 0.78. While this variant shows some improvement, it does not surpass the performance of the focal loss function variant.
- **SCANSleepNet (Ours):** The proposed SCANSleepNet, which includes IFFE, two TFC modules, and the WFCE loss function, achieved the highest performance with an accuracy of 85.52%, an MF1 score of 78.31, and a k value of 0.80. These results underscore the effectiveness of the WFCE loss function in addressing class imbalance and challenging samples.

Table 4 Comparative results of computational efficiency in Flops and Parameters

Model	# Flops (M)	# Parameters (M)
U-time	174.27	2.37
CCRR	2197.21	25.71
deepsleep	1050.96	22.65
sleepeeg	198.29	2.64
AttnSleep	61.97	0.72
Sleepeeg	499.44	24.08
SCAN(ours)	31.15	0.26

The comparative results are shown in Table 5. It demonstrates the significant contributions of each component within SCANSleepNet. Specifically, the TFC module plays a crucial role in the model. They capture time-related features of EEG data transitions between different stages, which significantly enhances the model's ability to learn complex patterns. Additionally, the Weighted focal cross-entropy loss function excels in addressing data imbalance issues. This function dynamically adjusts the weights of hard-to-classify samples and initially determines the penalty for each class based on the number of classes, ensuring that the model focuses more on minority and difficult-to-classify samples during training and avoids potential overfitting problems associated with traditional loss functions. Overall, SCANSleepNet demonstrates robust capabilities in accurately classifying sleep stages. Its deep learning of temporal features and effective handling of data imbalance highlight its effectiveness and potential for practical applications in sleep monitoring.

5.2 Variation of transform

In this section, we discussed Transform's feature extraction capabilities. In the IFFE module, we replaced the CNN feature extraction preceding the MSCA block with a Transformer encoder. Given computational resource constraints, the Transformer encoder was limited to 1 layer and the multi-head attention mechanism was configured with 4 heads. While this simplification partially mitigates computational resource constraints, it inevitably weakens the encoder's information representation capability. As shown in Table 6, the performance of the model declines significantly after the adoption of the Transformer encoder. On the Sleep-EDF Fpz Cz channel, the overall accuracy drops from 85.52% to 75.70%, the macro-average F1 score decreases from 78.31% to 67.00%, and Cohen's kappa score falls from 0.80 to 0.66.

5.3 Variation of one-dimensional convolution

In this section, we explored the role of causal convolution in the TFC module. Specifically, we replaced causal convolution with one-dimensional convolution and conducted

Table 5 Ablation study of counterparts on the Sleep-EDF-20 dataset's Fpz-Cz channel

Methods	Focal	CA	WFCE	Acc. (%)	MF1 (%)	k
IFFE			✓	84.02	76.73	0.78
IFFE + one TFC			✓	84.07	78.00	0.78
IFFE + two TFC	✓			84.60	77.39	0.79
IFFE + two TFC		✓		84.21	76.20	0.78
SCANSleepNet (ours)			✓	85.52	78.31	0.80

The symbol '✓' indicates which loss function is used in this method. The best and second-best results are marked in red and blue, respectively

experiments on the Sleep-EDF Fpz-Cz channel. As shown in Table 6, causal convolution is critical for capturing causal relationships and long-range dependencies in temporal signals. After replacing causal convolution with one-dimensional convolution, the overall accuracy declined to 84.56%, the Macro-F1 score decreased from 78.31% to 77.02%, and the Cohen's kappa score also dropped from 0.80 to 0.79. The performance decline stems from one-dimensional convolution's inability to explicitly model causal relationships in temporal data, limiting its capacity to capture complex temporal dynamics.

6 Conclusion

We propose a sleep stage classification architecture termed SCANSleepNet to categorize sleep stages using raw EEG signals from a single channel. A multiscale atrous convolution (MSAC) block is designed to derive features from EEG signals. Meanwhile, a Temporal Feature Capture (TFC) module is constructed to capture temporal correlation. Besides, a weighted focus cross-entropy loss function is formulated to manage class imbalance effectively and focus on difficult samples. Experiment results prove that it achieves a good balance between classification accuracy and efficiency. Given the fact that SCANSleepNet has demonstrated certain advantages in the task of sleep stage classification, there is still some room for improvement. First, the classification effect of the N1 stage is significantly lower than that of other stages. Second, the model structure of SCANSleepNet has not been optimized for the rapidly changing features of the N1 stage. Future work is expected on N1 sleep stage classification.

Table 6 Performance comparison between SCAN (Ours) and Transform variants and One-dimensional convolution variants

Methods	Acc.(%)	MF1(%)	k
SCAN(Ours)	85.52	78.31	0.80
Transform	75.70	67.00	0.66
One-dimensional convolution	84.56	77.02	0.79

Acknowledgements This work was funded by Shandong Province Undergraduate Teaching Reform Project (No.Z2024184).

Author Contributions Yuyun Liu: Conceptualization and Original draft. Qilei Li: Data curation and Formal analysis. Mingliang Gao: Supervision, Review and Editing. Xiangyu Guo: Methodology and English polishing. Wenzhe Zhai: Network development.

Data Availability The data that support the findings of this study are available from authors upon reasonable request.

Declarations

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

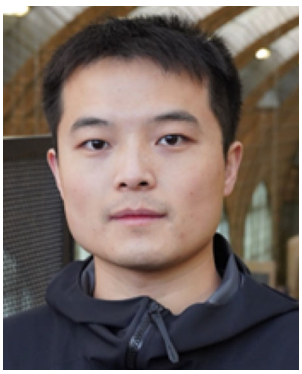
Ethical and Informed Consent for Data Used Not applicable.

References

1. Luyster FS, Strollo PJ Jr, Zee PC, Walsh JK (2012) Sleep: a health imperative. Sleep 35(6):727–734
2. Rauchs G, Desgranges B, Foret J, Eustache F (2005) The relationships between memory systems and sleep stages. J Sleep Res 14(2):123–140
3. Sharma S, Kavuru M et al (2010) Sleep and metabolism: an overview. Int J Endocrinol
4. Tank J, Diedrich A, Hale N, Niaz FE, Furlan R, Robertson RM, Mosqueda-Garcia R (2003) Relationship between blood pressure, sleep k-complexes, and muscle sympathetic nerve activity in humans. Am J Physiol Regul Integr Comp Physiol 285(1):R208–R214
5. Malafeev A, Laptev D, Bauer S, Omlin X, Wierzbicka A, Wichniak A, Jernajczyk W, Riener R, Buhmann J, Achermann P (2018) Automatic human sleep stage scoring using deep neural networks. Front Neurosci 12(100):781
6. Supratak A, Dong H, Wu C, Guo Y (2017) Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg. IEEE Trans Neural Syst Rehabil Eng 25(11):1998–2008
7. Keenan SA (2005) An overview of polysomnography. Handb Clin Neurophysiol 6:33–50
8. Memar P, Faradji F (2017) A novel multi-class eeg-based sleep stage classification system. IEEE Trans Neural Syst Rehabil Eng 26(1):84–95
9. Neng W, Lu J, Xu L (2021) Ccrrsleepnet: A hybrid relational inductive biases network for automatic sleep stage classification on raw single-channel eeg. Brain Sci 11(4):456

10. Mousavi S, Afghah F, Acharya UR (2019) Sleepeegnet: automated sleep stage scoring with sequence to sequence deep learning approach. PLoS ONE 14(5):e0216456
11. Zhao C, Li J, Guo Y (2024) Sequence signal reconstruction based multi-task deep learning for sleep staging on single-channel eeg. Biomed Signal Process Control 88:105615
12. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) Smote: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
13. Chen M, Gui Y, Su Y, Zhu Y, Luo G, Yang Y Improving eeg classification through randomly reassembling original and generated data with transformer-based diffusion models. arXiv:2407.20253
14. Ying S, Li P, Chen J, Cao W, Zhang H, Gao D, Liu T (2025) An eeg-based single-channel dual-stream automatic sleep staging network with transfer learning. Appl Soft Comput 170:112722
15. Wang Z, Liu Y, Huang J An open api architecture to discover the trustworthy explanation of cloud ai services. IEEE Trans Cloud Comput
16. Tsinalis O, Matthews PM, Guo Y, Zafeiriou S Automatic sleep stage scoring with single-channel eeg using convolutional neural networks. arXiv:1610.01683
17. Chriskos P, Frantzidis CA, Gkivogkli PT, Bamidis PD, Kourtidou-Papadeli C (2019) Automatic sleep staging employing convolutional neural networks and cortical connectivity images. IEEE Trans Neural Netw Learn Syst 31(1):113–123
18. Sokolovsky M, Guerrero F, Paisarnsrisomsuk S, Ruiz C, Alvarez SA (2019) Deep learning for automated feature discovery and classification of sleep stages. IEEE/ACM Trans Comput Biol Bioinform 17(6):1835–1845
19. Yin H, Mukadam B, Dai X, Jha NK (2019) Diabdeep: Pervasive diabetes diagnosis based on wearable medical sensors and efficient neural networks. IEEE Trans Emerg Top Comput 9(3):1139–1150
20. Xiao Z, Tong H, Qu R, Xing H, Luo S, Zhu Z, Song F, Feng L Capmatch: semi-supervised contrastive transformer capsule with feature-based knowledge distillation for human activity recognition. IEEE Trans Neural Netw Learn Syst
21. Xiao Z, Xing H, Qu R, Feng L, Luo S, Dai P, Zhao B, Dai Y (2024) Densely knowledge-aware network for multivariate time series classification. IEEE Trans Syst Man Cybern Syst 54(4):2192–2204
22. Yang S, Linares-Barranco B, Wu Y, Chen B Self-supervised high-order information bottleneck learning of spiking neural network for robust event-based optical flow estimation. IEEE Trans Pattern Anal Mach Intell
23. Yang S, Chen B Effective surrogate gradient learning with high-order information bottleneck for spike-based machine intelligence. IEEE Trans Neural Netw Learn Syst
24. Yang S, Chen B (2023) Snib: improving spike-based machine learning using nonlinear information bottleneck. IEEE Trans Syst Man Cybern Syst 53(12):7852–7863
25. Eldele E, Chen Z, Liu C, Wu M, Kwoh C-K, Li X, Guan C (2021) An attention-based deep learning approach for sleep stage classification with single-channel eeg. IEEE Trans Neural Syst Rehabil Eng 29:809–818
26. Wang Z, Liu Y Staa: spatio-temporal attention attribution for real-time interpreting transformer-based video models. arXiv:2411.00630
27. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pmlr, pp 448–456
28. Misra D Mish: a self regularized non-monotonic activation function. arXiv:1908.08681
29. Kingma DP, Ba J Adam: a method for stochastic optimization. arXiv:1412.6980
30. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220
31. Sun Y, Wang B, Jin J, Wang X (2018) Deep convolutional network method for automatic sleep stage classification based on neurophysiological signals, in: 2018 11th international Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, pp 1–5
32. Phan H, Andreotti F, Cooray N, Chén OY, De Vos M (2018) Joint classification and prediction cnn framework for automatic sleep stage classification. IEEE Trans Biomed Eng 66(5):1285–1296
33. Perslev M, Jensen M, Darkner S, Jennum PJ, Igel C U-time: a fully convolutional network for time series segmentation applied to sleep staging. Adv Neural Infor Process Syst 32

Yuyun Liu received the B.E. degree with the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include sleep stage classification and deep learning.



Qilei Li received the Ph.D. in Computer Science from Queen Mary University of London. He previously earned an M.S. degree from Sichuan University in 2020. From June 2022 to April 2024, he worked as a machine learning scientist at Veritone Inc, where he focused on developing a scalable person search framework for retrieving individuals at different locations and times, as captured by various cameras. His current research interests lie in privacy-aware machine learning, with a particular emphasis on learning domain-invariant knowledge representation from multimodal data captured in diverse environments. His research outcome has been recognized as ESI Highly Cited Paper (Top 1%). Additionally, he serves as an evaluator for the ELLIS PhD Program.

Mingliang Gao received his Ph.D. in Communication and Information Systems from Sichuan University. He is now an associate professor at the Shandong University of Technology. He was a visiting lecturer at the University of British Columbia during 2018-2019. He has been the principal investigator for a variety of research funding, including the National Natural Science Foundation, the China Postdoctoral Foundation, National Key Research Development Project, etc. His research interests include computer vision, machine learning, and intelligent optimal control. He has published over 150 journal/conference papers in IEEE, Springer, Elsevier, and Wiley. He serves as a reviewer for more than 30 journals, e.g., Information Fusion, IEEE Transaction on Image processing, Pattern recognition, and IEEE Transactions on Instrumentation & Measurement.

Xiangyu Guo received the M.S. degree with the School of Electrical and Electronic Engi neering, Shandong University of Technology, Zibo, China. His research interests include smart city systems, computer vision, and deep learning.

Wenzhe Zhai received the M.S. degree at the School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo, China. His research interests include smart city system, information fusion, crowd analysis and deep learning.