DilatedSleepNet: A Novel EEG Waveform-Aware Model for Single-Channel Automatic Sleep Staging

Zekun Zheng, Zhuorong Li, Pei Mei, Fei Wang*
School of Artificial Intelligence, South China Normal University, China

Abstract

Sleep plays a crucial role in maintaining human health and improving quality of life. However, traditional manual sleep staging methods are not only time-consuming but also heavily reliant on expert experience, limiting their feasibility for large-scale applications. Therefore, developing high-precision and fully automated sleep staging methods is essential for assisting clinical diagnosis. To address this research need, we propose an innovative automatic sleep staging network, DilatedSleepNet. This model introduces a novel multi-scale dilated convolution strategy to effectively capture the waveform characteristics of EEG signals, enabling accurate sleep stage classification using only single-channel EEG input. We systematically evaluated the performance of DilatedSleepNet on three publicly available datasets, achieving classification accuracies of 86.8%, 83.2%, and 85.4%, respectively. Experimental results demonstrate that DilatedSleepNet exhibits outstanding generalization ability and robustness across multiple datasets, providing a strong technical foundation for the diagnosis and research of sleep-related disorders.

Keywords: Sleep Staging; Dilated convolution; Single-channel EEG

Introduction

Sleep plays a vital role in maintaining overall health (Mason, Lokhandwala, Riggins, & Spencer, 2021). Insufficient sleep has been closely linked to an increased risk of chronic conditions such as hypertension, stroke, and diabetes (Vallat, Shah, Redline, Attia, & Walker, 2020). Clinically, sleep staging is commonly used as a key indicator for assessing sleep quality. Polysomnography (PSG) is regarded as the gold standard for sleep staging, integrating various measurement techniques, including electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), and electrocardiography (ECG) (Khalili & Asl, 2021). These modalities respectively record brain electrical activity, eye movements, muscle activity, and heart rate variability. Based on these signals and following the standard guidelines set by the American Academy of Sleep Medicine (AASM), sleep is classified into distinct stages. The AASM guidelines classify sleep into five stages: Wakefulness (W), Rapid Eye Movement (REM), and three Non-Rapid Eye Movement (NREM) stages—N1 (transition), N2 (light sleep), and N3 (deep sleep) (Jorgensen et al., 2020). Manual sleep staging is a time-consuming, labor-intensive, and subjective process (Fiorillo et al., 2019). Developing an accurate and reliable automatic sleep staging method can significantly enhance efficiency and assist sleep specialists in making diagnostic evaluations.

In recent years, deep learning has made significant progress in automatic sleep staging. Methods like TSC-AEA efficiently extract robust features from limited labeled data (Li, Chen, Pan, & Huang, 2024), while Deepsleepnet-Lite ensures computational efficiency and high accuracy, particularly for resource-constrained scenarios (Fiorillo, Favaro, & Faraci, 2021). SleepKD leverages knowledge distillation to enhance accuracy and generalization (Liang, Liu, Wang, Jia, & Center, 2023). Additionally, the exploration of Deep Belief Networks (DBN) (Längkvist, Karlsson, & Loutfi, 2012) and Deep Neural Networks (DNN) (Dong et al., 2017) has enriched technical approaches and broadened application prospects.

Single-channel automatic sleep staging has become a research focus due to its advantages in simplifying data collection, reducing computational demands, and lowering hardware costs. For instance, DeepSleepNet uses multi-layer CNNs to improve staging accuracy (Supratak, Dong, Wu, & Guo, 2017), AttnSleep employs attention mechanisms to enhance performance and interpretability (Eldele et al., 2021), and MRASleepNet adopts multi-task learning to boost accuracy and generalization (Yu, Zhou, Wu, Gao, & Bin, 2022). However, single-channel data face challenges in capturing key frequency information and transitions, limiting the identification of certain sleep stages.

To address these issues, we propose a novel single-channel EEG-based sleep staging method. By integrating tailored architectures and modules, it improves staging performance and overcomes limitations in capturing frequency information, demonstrating significant potential for practical applications. This study's main contributions are summarized as follows:

1) We present DilatedSleepNet, an automatic sleep staging method that demonstrates excellent performance using only single-channel EEG signals.

2) We introduce the Hierarchical Temporal Dilation Network and the Multi-Scale Temporal Dilation Integrator modules, which effectively capture the crucial frequency information and frequency transitions characteristic of the sleep process.

3) We evaluate our proposed model, DilatedSleepNet, on three publicly available datasets, and it exhibits exceptional performance in single-channel automatic sleep staging.

Methods

Model Overview

Figure 1 illustrates the overall architecture of DilatedSleepNet, which is composed of four key components: the feature extraction module, the Hierarchical Temporal Dilation Net-

---

*Corresponding author: Fei Wang.
Email addresses: zhengzekun@m.scnu.edu.cn(Zekun Zheng), r6039963@gmail.com(Zhuorong Li), 845840977@qq.com(Pei Mei), fwang@scnu.edu.cn(Fei Wang)

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

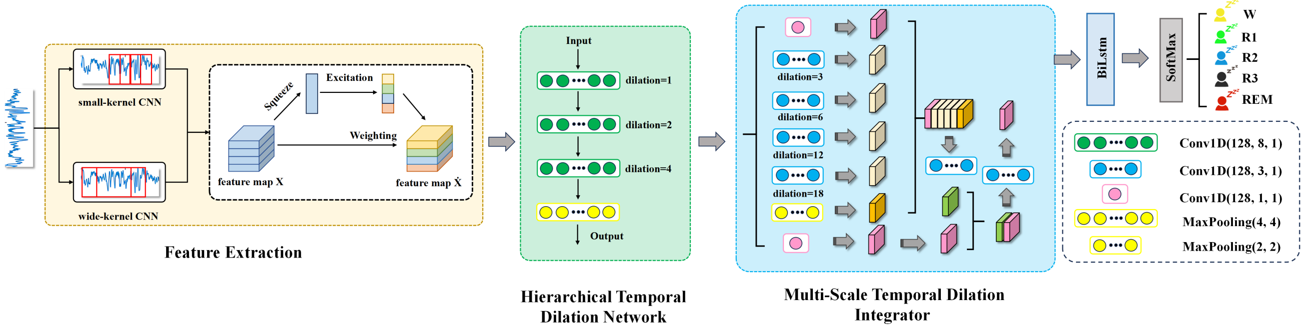


Figure 1: The overall architecture of the proposed DilatedSleepNet.

work (HTDN), the Multi-Scale Temporal Dilation Integrator (MTDI), and the temporal feature extraction module.

First, DilatedSleepNet leverages the feature extraction module to perform an initial analysis of the input single-channel EEG signals, extracting critical frequency characteristics. To comprehensively capture diverse frequency information, we adopted a dual-branch CNN structure inspired by prior research (Supratak et al., 2017). Additionally, we enhanced the traditional SE channel attention mechanism (Hu, Shen, & Sun, 2018), enabling more precise identification of critical features and improving the efficiency and effectiveness of feature extraction. Second, the Hierarchical Temporal Dilation Network (HTDN) is designed to model fine-grained temporal dependencies within time-series data. By employing a stacked dilated convolutional structure, this module excels in capturing transient but salient local signal patterns, such as sleep spindles and K-complexes (Caporro et al., 2012), which are often localized in specific time intervals and play a pivotal role in differentiating sleep stages. Through this framework, the model achieves enhanced capability in characterizing signal details. Third, the Multi-Scale Temporal Dilation Integrator (MTDI) focuses on modeling global temporal context across multiple time scales. The frequency characteristics of the sleep stage often exhibit dependencies that span various time scales (McCausland, Biglarbeigi, Bond, Yadollahikhales, & Finlay, 2022). MTDI employs a parallel receptive field design to integrate features from different temporal ranges, enabling a balanced representation of both global and local patterns. This design significantly enhances the model's capacity for comprehensive signal modeling. Lastly, the model incorporates a bidirectional long short-term memory (BiLSTM) network to further refine long-term temporal dependencies. By processing sequences in both forward and backward directions, BiLSTM effectively captures the dynamic characteristics of EEG signals, providing a more holistic and reliable representation of features for the final classification task. In subsequent sections, we will provide a detailed explanation of these modules.

Table 1: Parameters of the feature extraction.

Layer Name	Layer Type	Filters	Size	Stride	Output Size
Conv11	1d-Conv	128	50	6	(128, 1501)
Conv21	1d-Conv	128	400	50	(128, 181)
Pool11	1d-MaxPool	-	8	2	(128, 751)
Pool21	1d-MaxPool	-	4	2	(128, 91)
Dp11	Dropout0.5	-	-	-	(128, 752)
Dp21	Dropout0.5	-	-	-	(128, 92)
Conv12	1d-Conv	128	8	1	(128, 753)
Conv22	1d-Conv	128	6	1	(128, 93)
Conv13	1d-Conv	128	8	1	(128, 753)
Conv23	1d-Conv	128	6	1	(128, 93)
Pool12	1d-MaxPool	-	4	4	(128, 189)
Pool12	1d-MaxPool	-	2	2	(128, 47)
Dp12	Dropout0.5	-	-	-	(128, 189)
Dp22	Dropout0.5	-	-	-	(128, 47)
Cat	Concatnate	-	-	-	(128, 236)

Feature Extraction

To extract features spanning diverse frequency ranges, we devised a dual-domain convolutional network architecture, as illustrated in Table 1. This design is motivated by the distinctive frequency-domain characteristics inherent to various sleep stages, emphasizing the importance of selectively extracting features from specific frequency bands to optimize sleep staging accuracy(Memar & Faradji, 2017).

The first branch, utilizing wide convolutional kernels (*kernel size* = 400), is tailored for capturing low-frequency signal patterns. In a dataset sampled at 100 Hz, this branch processes approximately 4 seconds of temporal data per convolutional window, enabling it to capture the complete cycle of low-frequency signals. Such low-frequency elements, typically linked to the infra-slow oscillation band ($<0.5\ Hz$), are critical for recognizing deep sleep stages. Conversely, the second branch leverages smaller convolutional kernels (*ker-*

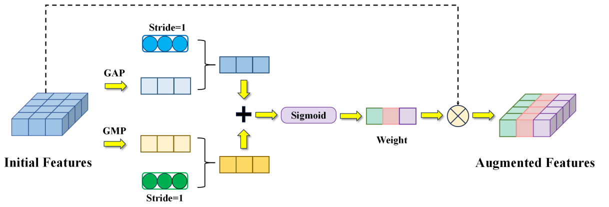


Figure 2: The structure of the enhanced channel attention mechanism

nel size = 50) to emphasize localized, higher-frequency features. Each convolutional window spans 0.5 seconds, facilitating the detection of signals with higher frequencies. This branch is adept at extracting features corresponding to α and θ waves, which are essential markers of rapid eye movement (REM) and light sleep stages.

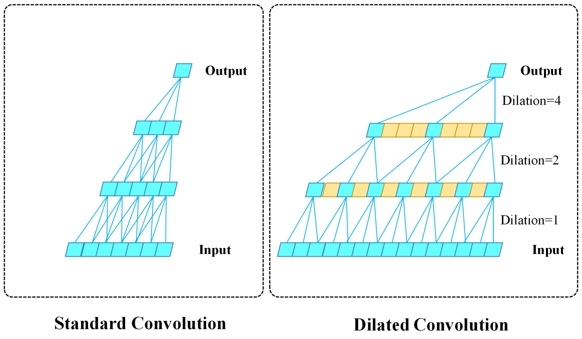
To enhance training stability and improve the expressive power of the neural network, batch normalization layers are applied after each convolutional layer(Ioffe, 2015), with the Gaussian Error Linear Unit (GELU) chosen as the activation function(Hendrycks & Gimpel, 2016). By fusing the outputs from both branches, the dual-domain convolutional network integrates low-frequency global patterns with detailed high-frequency local features, delivering a holistic representation of the physiological dynamics characterizing different sleep stages.

To improve feature representation, this study refines the SE channel attention mechanism to enhance the selective representation of channel features. The updated approach integrates both max pooling and average pooling, with max pooling used to capture significant features and average pooling offering more generalized, global information, as illustrated in Figure 2. This refinement allows the network to selectively emphasize or suppress certain channels, thereby enhancing feature representation for subsequent tasks.

Hierarchical Temporal Dilation Network

EEG signals, such as Delta waves and spindle waves, display significant temporal dependencies and long-term structural patterns. However, traditional convolutional networks are constrained by their receptive fields, which primarily capture local features within limited time frames. This limitation hinders their ability to effectively capture the global patterns and intricate details of these signals.

To overcome this challenge, we proposed the Hierarchical Temporal Dilation Network (HTDN), which gradually increases the dilation rate to expand the receptive field. This approach enhances the network's capacity to capture long-range dependencies in time-series data. Figure 3 compares the receptive fields of standard convolution and HTDN with a kernel size of 3. Despite using the same number of parameters, HTDN's receptive field is expanded by more than twice compared to that of traditional convolution. In the Dilated-SleepNet framework, we use a kernel size of 8. The larger kernel size enables the network to cover a wider local feature space, significantly improving its ability to model both low-frequency and high-frequency signals, while maintaining fine-grained feature resolution.



Standard Convolution

Dilated Convolution

Figure 3: The comparison of receptive fields in standard and dilated convolutions

The receptive field in HTDN is calculated as follows:

$$R = \sum_{i=1}^{n} \left((K-1) \cdot \text{dilation}_i + 1 \right) \tag{1}$$

where R represents the receptive field, K is the size of the convolution kernel, and dilation_i is the dilation rate at the i-th layer. The total receptive field is obtained by summing the contributions of all layers in the dilated convolution stack.

Multi-Scale Temporal Dilation Integrator

In the processing of sleep signals, different sleep waveforms exhibit significant variations in their temporal cycles and spans. To address this issue, we propose the Multi-Scale Temporal Dilation Integrator(MTDI) module, which leverages multiple convolutional branches with varying dilation rates to capture features at different temporal scales, thereby enhancing the model's ability to process signals spanning diverse time ranges.

The MTDI module consists of several convolutional branches, each configured with a distinct dilation rate r. Smaller dilation rates focus on capturing short-term local features, while larger dilation rates are designed to model long-term global dependencies. The dilation rates are set as $r \in \{1, 3, 6, 12, 18\}$. Given an input signal $x \in \mathbb{R}^{C \times T}$, where C is the number of channels and T is the temporal length, the output of each branch i can be expressed as:

$$y_i = \text{GELU}(\text{BN}(\text{Conv1D}(x, r_i))), \tag{2}$$

where $\text{Conv1D}(x, r_i)$ represents the 1D convolution operation with a dilation rate r_i, BN denotes batch normalization, and GELU is the activation function.

Additionally, a global average pooling (GAP) branch is included to extract global information from the signal. The global characteristic g is defined as:

$$g = \text{Interpolate}(\text{GAP}(x), T), \tag{3}$$

where $\text{GAP}(x)$ performs the global temporal average pooling and Interpolate adjusts the global feature to match the temporal length T of the input. The outputs from all branches are concatenated and further processed through convolution to integrate multi-scale features effectively.

Experiments

Datasets

In this study, we utilize three widely used publicly available sleep datasets—SleepEDF-20, SleepEDF-78, and the Sleep Heart Health Study (SHHS)—to evaluate the performance of DilatedSleepNet. The details of each dataset are provided in Table 6. It is important to note that only single-channel data from each dataset were selected for the experiments.

Table 2: Details of the employed datasets.

Datasets	W	N1	N2	N3	REM	Total
SleepEDF-20	8285	2804	17799	5703	7717	42308
SleepEDF-78	65951	21522	69132	13039	25835	195479
SHHS	46319	10304	142125	60153	65953	324854

1) SleepEDF-20: The SleepEDF-20 dataset examines the relationship between age and sleep characteristics in healthy individuals aged 25 to 101 years (Kemp, Zwinderman, Tuk, Kamphuisen, & Oberye, 2000) (Rechtschaffen, 1968). It includes PSG recordings from 20 participants (10 males, 10 females), with one night missing for Subject 13 due to technical issues. The recordings, sampled at 100 Hz, include EEG and EOG signals, with each 30-second epoch classified into eight stages by sleep specialists using the R&K standard.

2) SleepEDF-78: This dataset extends SleepEDF-20, consisting of PSG recordings from 78 participants aged 25 to 101 years (Goldberger et al., 2000). Each participant had two consecutive nights of PSG data, with EEG and EOG signals sampled at 100 Hz. Sleep stages were classified into eight categories by specialists following the R&K criteria. One night of recordings is missing for participants 13, 36, and 52 due to equipment malfunctions.

3) SHHS: The SHHS dataset, a large-scale collection from multiple medical centers, examines the relationship between sleep disorders and cardiovascular diseases (Quan et al., 1997) (Zhang et al., 2018). For this study, 329 participants with an apnea-hypopnea index (AHI) of less than 5 and regular sleep patterns were selected. Sleep stages were classified into eight categories by sleep experts using the R&K standard.

Experiment Details

In the SleepEDF-20 and SleepEDF-78 datasets, EEG signals from the Fpz-Cz channel sampled at 100 Hz were utilized, while for the SHHS dataset, signals from the C4-A1 channel sampled at 125 Hz were selected. To ensure consistency and standardization, the ”Unknown” and ”Movement” categories were excluded, and the N3 and N4 stages were merged following the AASM guidelines. Additionally, to comprehensively capture transitions in sleep states, 30 minutes of data preceding and following the sleep period were retained during preprocessing.

Table 3: Confusion matrix from 20-Fold cross-Validation on SleepEDF-20 Fpz-Cz channel.

	Predicted					Per-class metrics		
	W	N1	N2	N3	REM	PR	RE	F1
W	7565	408	174	29	119	90.8	90.9	90.8
N1	295	1417	634	16	442	56.5	50.5	53.3
N2	294	369	15875	710	551	89.7	89.2	89.4
N3	22	4	399	5276	2	87.4	92.5	89.9
REM	153	302	612	7	6643	85.5	86.1	85.8

Table 4: Confusion matrix from 10-Fold cross-validation on SleepEDF-78 Fpz-Cz channel.

	Predicted					Per-class metrics		
	W	N1	N2	N3	REM	PR	RE	F1
W	62109	2703	399	58	682	91.2	94.2	92.7
N1	4385	9094	6101	137	1805	53.3	42.3	47.2
N2	658	3540	59328	710	551	85.1	85.8	85.5
N3	38	13	1594	11367	27	77.0	87.2	81.8
REM	891	1699	2304	133	20808	80.5	80.5	80.5

Table 5: Confusion matrix from 20-Fold cross-validation on SHHS C4-A1 channel.

	Predicted					Per-class metrics		
	W	N1	N2	N3	REM	PR	RE	F1
W	40916	1883	1903	326	1307	84.0	88.3	86.1
N1	1442	4975	1900	14	1973	47.1	48.3	47.7
N2	3761	2389	121375	8125	6475	88.8	85.4	87.0
N3	701	12	8530	50544	366	85.3	84.0	84.7
REM	1876	1297	2985	219	59576	85.5	90.3	87.8

Inspired by previous studies, we employ a 90-second epoch (Z_m) as the contextual input to the model. This input comprises three consecutive 30-second epochs: the preceding epoch (X_{m-1}), the current epoch (X_m), and the succeeding epoch (X_{m+1}). The ground truth label of Z_m is y_m, which also corresponds to the label of X_m. This relationship can be expressed as:

$$Z_m = (X_{m-1}, X_m, X_{m+1}) \mapsto y_m \tag{4}$$

Such a design enables the model to capture the dependencies and correlations among epochs. This approach aligns with the procedure adopted by sleep specialists, who rely not only on the features of a single epoch but also on temporal characteristics of adjacent epochs to make a more informed and accurate sleep stage classification.

To ensure model generalization across participants and

Table 6: Comparison between DilatedSleepNet (our method) and other methods

Dataset	Method	Subjects	Per-Class F1-score					Overall Metrics		
			W	N1	N2	N3	REM	Acc.	MF1	κ
SleepEDF-20	DeepSleepNet	20	84.7	46.6	85.9	84.8	82.4	82.0	76.9	0.76
	SleepEEGNet	20	89.2	52.2	86.8	85.1	85.0	84.3	79.7	0.79
	AttnSleep	20	89.7	42.6	88.8	**90.2**	79.0	84.4	78.1	0.79
	MRASleepNet	20	90.7	46.1	88.7	88.2	80.7	84.5	78.9	0.79
	SsleepNet	20	89.5	50.2	87.6	87.0	83.0	84.6	79.5	-
	DilatedSleepNet	20	**90.8**	**53.3**	**89.4**	89.9	**85.8**	**86.8**	**81.9**	**0.82**
SleepEDF-78	DeepSleepNet	78	90.8	44.8	78.5	67.9	71.3	76.9	70.7	0.69
	SleepEEGNet	78	91.7	44.1	82.5	73.5	75.2	80.0	73.6	0.73
	MRASleepNet	78	92.0	44.5	84.9	80.1	76.1	81.4	75.4	0.74
	AttnSleep	78	92.0	42.0	85.0	**82.1**	74.2	81.3	75.1	0.74
	DilatedSleepNet	78	**92.7**	**47.2**	**85.5**	81.8	**80.5**	**83.2**	**77.5**	**0.77**
SHHS	DeepSleepNet	329	85.4	40.5	82.5	79.3	81.9	81.0	73.9	0.73
	SleepEEGNet	329	81.3	34.4	73.4	75.9	77.0	73.9	68.4	0.65
	AttnSleep	329	**86.7**	33.2	**87.1**	**87.1**	82.1	84.2	75.3	0.78
	SsleepNet	329	86.0	**47.8**	85.6	82.0	86.3	84.3	77.5	-
	DilatedSleepNet	329	86.1	47.7	**87.1**	84.7	**87.8**	**85.4**	**78.7**	**0.80**

prevent data leakage, we employed cross-validation techniques for evaluating the performance of our model on the SleepEDF-20, SleepEDF-78, and SHHS datasets. Specifically, we applied 20-fold cross-validation on SleepEDF-20, 10-fold cross-validation on SleepEDF-78, and 20-fold cross-validation on SHHS. In the case of the SleepEDF-20 dataset, consisting of data from 20 participants, each participant's data was iteratively used as the validation set, with the remaining participants' data serving as the training set. This strategy guarantees that every participant's data is used for validation exactly once, while avoiding any overlap with the training set. Finally, we aggregated the predicted sleep stages from all folds to compute the overall performance metrics, thereby ensuring a comprehensive evaluation of the model's accuracy and reliability. We utilized a batch size of 128 for training, employing the Adam optimizer with an initial learning rate of 4e-4, which was reduced to 4e-5 after the first 10 epochs. Training spanned 100 epochs for each fold during cross-validation. The Adam optimizer was configured with the following parameters: a weight decay of 1e-3, β_1 and β_2 values of 0.9 and 0.999, respectively, an epsilon value of 1e-8, and AMSGrad enabled.

Baselines

In this experiment, we compared the proposed model with five baseline models. A brief introduction to each baseline is provided below:

1) DeepSleepNet: Combines a dual-layer CNN and LSTM networks to extract features for automatic sleep stage classification (Supratak et al., 2017).

2) SleepEEGNet: A lightweight CNN designed to efficiently process single-channel EEG signals for sleep staging (Mousavi, Afghah, & Acharya, 2019).

3) AttnSleep: Incorporates multi-head attention mechanisms to capture temporal dependencies and key features in sleep stages (Eldele et al., 2021).

4) MRASleepNet: Utilizes multi-resolution analysis and attention mechanisms to improve the modeling of multi-band EEG features (Yu et al., 2022).

5) SsleepNet: Integrates temporal characteristics of signals with deep learning frameworks to enhance sleep staging performance for single-channel EEG data (Lv, Ma, Li, & Ren, 2024).

Performance Comparison

Tables 3, 4 and 5 present the confusion matrices of the proposed model evaluated on the SleepEDF-20 and SleepEDF-78 datasets using the FPZ-CZ channel with 20-fold and 10-fold cross-validation, as well as on the SHHS dataset using the C4-A1 channel with 20-fold cross-validation. Each row represents the number of samples labeled by experts, while each column denotes the number of samples predicted by the model for each sleep stage. Overall, the N1 stage, due to its relatively small sample size, is often misclassified as W, REM, or N2, resulting in an F1 score for the N1 stage that is typically below 50%.

Table 6 presents a performance comparison between DilatedSleepNet and five state-of-the-art baseline models (DeepSleepNet, SleepEEGNet, AttnSleep, MRASleepNet, and SSleepNet), offering a comprehensive evaluation of their effectiveness in automatic sleep staging. The experiments were conducted on the SleepEDF-20, SleepEDF-78, and SHHS datasets, employing metrics such as accuracy (ACC), macro-averaged F1 score (MF1), and Cohen's kappa coefficient (κ) to systematically analyze the classification capabilities of each model.

On the SleepEDF-20 dataset, DilatedSleepNet demonstrated outstanding performance, achieving an accuracy of 86.8%, an MF1 score of 81.9%, and a kappa coefficient of 0.82, surpassing all baseline models. On the SleepEDF-78 dataset, the model further validated its generalization capability, with accuracy, MF1, and kappa reaching 83.2%, 77.5%, and 0.77, respectively, maintaining robust performance. For the larger-scale SHHS dataset, DilatedSleepNet continued to exhibit exceptional adaptability and generalization, underscoring its ability to handle heterogeneous and large-scale data effectively. These results indicate that DilatedSleepNet excels not only in small-scale datasets but also in complex, large-scale sleep staging scenarios.

DilatedSleepNet particularly excelled in classifying challenging sleep stages, such as REM, N1, and N2, which are characterized by high-frequency, low-amplitude signals often accompanied by specific features such as sleep spindles. By leveraging the HDTN and MTDI modules, the model effectively extracted and processed these intricate signal characteristics, significantly enhancing classification accuracy. However, its performance in the N3 stage, which is dominated by low-frequency slow waves with fewer high-frequency, low-amplitude features, was comparatively modest.

In summary, DilatedSleepNet demonstrates superior performance in capturing multi-band features, handling complex signals, and adapting to datasets of varying scales and heterogeneity. Its remarkable overall accuracy and strong performance in challenging sleep stages highlight its potential for clinical applications, offering a reliable tool for sleep data analysis and improving the efficiency and accuracy of sleep disorder diagnosis.

Ablation Study

In this study, we introduce DilatedSleepNet, a novel architecture that integrates the HDTN, MDTI, and BiLSTM modules into the feature extraction network. To evaluate the contribution of each module to the overall performance of DilatedSleepNet, we conducted an ablation study on the SleepEDF20 dataset. The study involved systematically removing one module at a time, resulting in three distinct model variants:

1) DilatedSleepNet-HTDN: Incorporates the feature extraction module, MDTI, and BiLSTM, while excluding the HTDN module.

2) DilatedSleepNet-MTDI: Incorporates the feature extraction module, HDTN, and BiLSTM, while excluding the MTDI module.

3) DilatedSleepNet-BiLSTM: Incorporates the feature extraction module, HDTN, and MDTI, while excluding the BiLSTM module.

4) DilatedSleepNet: Fully integrates the HTDN, MTDI, and BiLSTM modules into the feature extraction network.

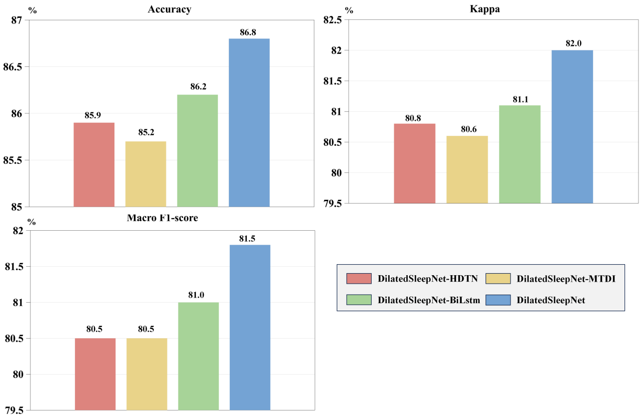


Figure 4: Ablation study conducted on SleepEDF-20 dataset.

Figure 4 presents the results of our ablation study, focusing on three key metrics: Accuracy, Kappa, and Macro F1-score. First, the ablation experiments confirm that all three integrated modules are essential for enhancing the model's classification performance. Second, comparisons between each variant and DilatedSleepNet demonstrate that both the HTDN and MTDI modules significantly contribute to performance improvement, with their absence leading to a 1%-2% reduction in overall performance. Additionally, the BiLSTM module also provides benefits by capturing temporal relationships among features. Finally, DilatedSleepNet outperforms all three variants across the three evaluation metrics, validating the effectiveness and rationality of its architectural design.

Conclusion

In this study, we introduce an innovative network architecture named DilatedSleepNet, designed for automatic sleep stage classification using single-channel EEG signals. Comprehensive evaluation on the publicly available datasets SleepEDF-20, SleepEDF-78, and SHHS demonstrates that DilatedSleepNet outperforms existing state-of-the-art methods across multiple key evaluation metrics, showcasing its exceptional classification performance and broad applicability.

Furthermore, through ablation experiments, we systematically analyze the contributions of the HTDN, MTDI, and BiLSTM modules to the model's performance. Specifically, the HTDN and MTDI modules play crucial roles in capturing multi-frequency features and improving classification accuracy, while the BiLSTM module further enhances the model's overall representational power by modeling temporal relationships. Experimental results show that DilatedSleepNet significantly outperforms its variants in terms of Accuracy, Kappa, and Macro F1-score, providing strong evidence for the soundness and efficiency of its architectural design.

Future research will focus on exploring the application of transfer learning techniques, particularly in transferring the models trained on labeled datasets to sleep stage classification tasks in patients with functional sleep disorders.

Acknowledgments

This work was supported in part by the STI 2030—Major Projects under Grant 2022ZD0208900, and in part by the Special Innovation Projects of Colleges and Universities in Guangdong Province under Grant 2023KTSCX030.

References

Caporro, M., Haneef, Z., Yeh, H. J., Lenartowicz, A., Buttinelli, C., Parvizi, J., & Stern, J. M. (2012). Functional mri of sleep spindles and k-complexes. *Clinical neurophysiology*, *123*(2), 303–309.

Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P. M., & Guo, Y. (2017). Mixed neural network approach for temporal sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(2), 324–333.

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 Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 809–818.

Fiorillo, L., Favaro, P., & Faraci, F. D. (2021). Deepsleepnet-lite: A simplified automatic sleep stage scoring model with uncertainty estimates. *IEEE transactions on neural systems and rehabilitation engineering*, *29*, 2076–2085.

Fiorillo, L., Puiatti, A., Papandrea, M., Ratti, P.-L., Favaro, P., Roth, C., … Faraci, F. D. (2019). Automated sleep scoring: A review of the latest approaches. *Sleep medicine reviews*, *48*, 101204.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., … Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, *101*(23), e215–e220.

Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7132–7141).

Ioffe, S. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Jorgensen, G., Downey, C., Goldin, J., Melehan, K., Rochford, P., & Ruehland, W. (2020). An australasian commentary on the aasm manual for the scoring of sleep and associated events. *Sleep and Biological Rhythms*, *18*, 163–185.

Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A., & Oberye, J. J. (2000). Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg. *IEEE Transactions on Biomedical Engineering*, *47*(9), 1185–1194.

Khalili, E., & Asl, B. M. (2021). Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel eeg. *Computer Methods and Programs in Biomedicine*, *204*, 106063.

Längkvist, M., Karlsson, L., & Loutfi, A. (2012). Sleep stage classification using unsupervised feature learning. *Advances in Artificial Neural Systems*, *2012*(1), 107046.

Li, J., Chen, Q., Pan, J., & Huang, H. (2024). A novel self-supervised learning method for sleep staging and its pilot study on patients with disorder of consciousness. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Liang, H., Liu, Y., Wang, H., Jia, Z., & Center, B. (2023). Teacher assistant-based knowledge distillation extracting multi-level features on single channel sleep eeg. In *Ijcai* (pp. 3948–3956).

Lv, X., Ma, J., Li, J., & Ren, Q. (2024). Ssleepnet: a structured sleep network for sleep staging based on sleep apnea severity. *Complex & Intelligent Systems*, *10*(2), 2689–2701.

Mason, G. M., Lokhandwala, S., Riggins, T., & Spencer, R. M. (2021). Sleep and human cognitive development. *Sleep medicine reviews*, *57*, 101472.

McCausland, C., Biglarbeigi, P., Bond, R., Yadollahikhales, G., & Finlay, D. (2022). Time-frequency ridge analysis of sleep stage transitions. In *2022 ieee signal processing in medicine and biology symposium (spmb)* (pp. 1–5).

Memar, P., & Faradji, F. (2017). A novel multi-class eeg-based sleep stage classification system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(1), 84–95.

Mousavi, S., Afghah, F., & Acharya, U. R. (2019). Sleepeegnet: Automated sleep stage scoring with sequence to sequence deep learning approach. *PloS one*, *14*(5), e0216456.

Quan, S. F., Howard, B. V., Iber, C., Kiley, J. P., Nieto, F. J., O'Connor, G. T., … others (1997). The sleep heart health study: design, rationale, and methods. *Sleep*, *20*(12), 1077–1085.

Rechtschaffen, A. (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. , 1–55.

Supratak, A., Dong, H., Wu, C., & Guo, Y. (2017). Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE transactions on neural systems and rehabilitation engineering*, *25*(11), 1998–2008.

Vallat, R., Shah, V. D., Redline, S., Attia, P., & Walker, M. P. (2020). Broken sleep predicts hardened blood vessels. *PLoS biology*, *18*(6), e3000726.

Yu, R., Zhou, Z., Wu, S., Gao, X., & Bin, G. (2022). Mrasleepnet: a multi-resolution attention network for sleep stage classification using single-channel eeg. *Journal of Neural Engineering*, *19*(6), 066025.

Zhang, G.-Q., Cui, L., Mueller, R., Tao, S., Kim, M., Rueschman, M., … Redline, S. (2018). The national sleep research resource: towards a sleep data commons. *Journal* *of the American Medical Informatics Association*, *25*(10), 1351–1358.