

Article

ADG-SleepNet: A Symmetry-Aware Multi-Scale Dilation-Gated Temporal Convolutional Network with Adaptive Attention for EEG-Based Sleep Staging

Hai Sun and Zhanfang Zhao * 

School of Information Engineering, Hebei GEO University, Shijiazhuang 052160, China; 18330054605@163.com

* Correspondence: zhaozhanfang@hgu.edu.cn

Abstract

The increasing demand for portable health monitoring has highlighted the need for automated sleep staging systems that are both accurate and computationally efficient. However, most existing deep learning models for electroencephalogram (EEG)-based sleep staging suffer from parameter redundancy, fixed dilation rates, and limited generalization, restricting their applicability in real-time and resource-constrained scenarios. In this paper, we propose ADG-SleepNet, a novel lightweight symmetry-aware multi-scale dilation-gated temporal convolutional network enhanced with adaptive attention mechanisms for EEG-based sleep staging. ADG-SleepNet features a structurally symmetric, parallel multi-branch architecture utilizing various dilation rates to comprehensively capture multi-scale temporal patterns in EEG signals. The integration of adaptive gating and channel attention mechanisms enables the network to dynamically adjust the contribution of each branch based on input characteristics, effectively breaking architectural symmetry when necessary to prioritize the most discriminative features. Experimental results on the Sleep-EDF-20 and Sleep-EDF-78 datasets demonstrate that ADG-SleepNet achieves accuracy rates of 87.1% and 85.1%, and macro F1 scores of 84.0% and 81.1%, respectively, outperforming several state-of-the-art lightweight models. These findings highlight the strong generalization ability and practical potential of ADG-SleepNet for EEG-based health monitoring applications.

Keywords: sleep staging; lightweight model; dilation-gated temporal convolutional network; multi-scale feature extraction; adaptive attention



Received: 11 June 2025

Revised: 7 August 2025

Accepted: 8 August 2025

Published: 5 September 2025

Citation: Sun, H.; Zhao, Z. ADG-SleepNet: A Symmetry-Aware Multi-Scale Dilation-Gated Temporal Convolutional Network with Adaptive Attention for EEG-Based Sleep Staging. *Symmetry* **2025**, *17*, 1461. <https://doi.org/10.3390/sym17091461>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sleep is a vital physiological process necessary for maintaining both physical and mental health. Scientific evaluation of sleep structure plays a key role in diagnosing sleep disorders, assessing disease risk, and guiding clinical treatments [1]. Traditionally, sleep staging is conducted by experts through manual analysis of polysomnography (PSG) recordings, following the standards set by the American Academy of Sleep Medicine (AASM). This process involves interpreting multi-channel and multi-modal signals such as electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG). However, this manual approach is time-consuming, labor-intensive, and inherently subjective [2], limiting its scalability in large-scale health screenings and routine clinical practices.

The burgeoning adoption of wearable health technologies and the accelerated growth of home-based medical monitoring have created pressing needs for resource-efficient sleep staging frameworks capable of automated real-time analysis [3,4]. Contemporary methodologies employ advanced computational approaches to derive discriminative features from

neurophysiological recordings (e.g., EEG), facilitating precise algorithmic categorization of sleep architecture into five distinct states (Wake, N1-N3, REM) [5,6]. This domain has witnessed transformative advancements following the integration of deep neural architectures [7,8].

Initial computational solutions for sleep stage classification predominantly utilized manually engineered features derived through temporal, spectral, or time-frequency analyses, subsequently processed by classical machine learning paradigms including support vector machines and ensemble decision trees [9,10]. These conventional approaches nevertheless demonstrated constrained cross-domain adaptability due to the inherent non-stationarity and inter-individual variability characteristic of EEG manifestations during sleep [11]. The advent of convolutional neural networks (CNNs) in 2016 revolutionized feature abstraction capabilities for this task [12,13]. Seminal work by Supratak et al. [7] introduced DeepSleepNet, which pioneered the synergistic integration of CNNs with bidirectional recurrent networks (BiLSTM), achieving unprecedented classification performance. Subsequent research has focused on architectural refinements of convolutional operators and temporal dependency modeling mechanisms to better accommodate the electrophysiological heterogeneity present in polysomnographic data [8,12,13].

The progressive deepening of neural network architectures has precipitated a corresponding surge in parametric complexity and computational overhead, creating significant barriers for resource-constrained implementations on mobile and embedded health monitoring platforms. In response to these constraints, the research community has investigated multiple optimization approaches: (1) compact convolutional topologies (e.g., depth-wise separable architectures [14], fire modules [15]), (2) network sparsification and precision reduction techniques [16], and (3) teacher-student paradigm distillation [17]. While effective in model compression, these methods frequently impair the hierarchical temporal representation capacity, ultimately diminishing model fidelity and cross-domain adaptability.

Clinical polysomnographic analysis reveals that EEG manifestations contain inherently multi-resolution temporal signatures—exemplified by delta-wave predominance in N3 sleep versus theta/gamma interplay during REM—necessitating sophisticated multi-scale feature aggregation for optimal staging performance [18,19]. Temporal convolutional architectures employing dilated kernels have demonstrated particular efficacy in establishing extended contextual awareness across divergent time scales [20]. Seminal implementations include Chambon et al.'s [13] stacked dilation framework for automated feature abstraction and Tsinalis et al.'s [12] systematic investigation of dilation coefficient impacts on stage-specific pattern recognition.

Contemporary TCN implementations nevertheless exhibit two critical limitations: (1) predetermined dilation configurations lacking input-adaptive flexibility [13], and (2) suboptimal local-global feature synthesis during transient sleep stage boundaries. Recent innovations attempt to mitigate these through hybrid multi-branch designs (e.g., MSTCN's parallelized multi-rate pathways [20]) or gated feature modulation mechanisms [21]. While MS-TCN [20] demonstrates improved multi-scale processing through static branch concatenation, they remain constrained by fixed fusion coefficients that prevent dynamic, context-dependent feature reweighting during temporal evolution.

Attention mechanisms have also been widely adopted in sequence modeling tasks, significantly enhancing discriminative capability by focusing on salient input features [22,23]. In automatic sleep staging, channel attention mechanisms such as Squeeze-and-Excitation (SE) blocks [24] and multi-head self-attention (MHA) modules [23] have increased sensitivity to critical features. Moreover, combining attention with dilated convolutions enables adaptive multi-scale feature aggregation [25].

Furthermore, gating mechanisms dynamically regulate information flow among parallel branches. While units such as the Gated Linear Unit (GLU) [26] and gated residual networks [27] are often task-specific, few studies have designed gating mechanisms explicitly for the dynamic fusion of different dilation rates. For example, Chen et al. [28] combined gating with multi-scale branches, but without channel attention or end-to-end adaptive feature selection. Effectively integrating parallel multi-dilated branches, channel attention, and adaptive gating for dynamic multi-scale feature fusion thus remains an open challenge.

Recently, research has increasingly focused on developing efficient, lightweight, and highly generalizable models for sleep staging. Notable examples include AttnSleep [29], which uses multi-head self-attention for long-range dependency modeling, and MSTCN [20] and NAMRTNet [30], which combine multi-scale TCN structures with channel attention. However, some of these methods involve complex architectures with large parameter counts, leading to high deployment costs, while others rely on fixed fusion weights and cannot dynamically adapt branch contributions to input data.

With the growing adoption of single-channel EEG in portable sleep monitoring, there is a pressing need for models that combine high accuracy, compact size, strong generalization, and efficient deployment [31,32]. To address these challenges, we propose ADG-SleepNet—a novel lightweight neural network architecture that integrates parallel multi-dilated convolutional branches, an adaptive gating mechanism, and channel attention. Our model extracts multi-scale temporal features in parallel and adaptively allocates branch importance based on input signals, while channel attention further enhances feature selection. This streamlined design enables significant reductions in parameter count and computational cost, supporting efficient deployment in resource-constrained environments. Experimental results demonstrate that ADG-SleepNet achieves state-of-the-art performance on public datasets, matching or exceeding leading lightweight models [7,32] and exhibiting strong generalization and application potential [5].

2. Materials and Methods

This section provides a comprehensive description of the materials and methodologies employed in this study. We first introduce the datasets used and the preprocessing procedures applied to the raw EEG signals. Subsequently, we detail the architecture and core components of the proposed ADG-SleepNet model.

2.1. Dataset and Preprocessing

To ensure the scientific rigor and broad applicability of model evaluation, this study utilizes the internationally recognized Sleep-EDF Database Expanded [33,34]. Specifically, two subsets of this database are employed for systematic model training, validation, and generalization testing. In addition, all raw EEG signals undergo preprocessing steps, including signal selection and the removal of irrelevant phases, prior to being input into the model. These procedures are essential for maximizing data quality and enhancing the stability of model training. The following sections will provide a detailed description of the dataset composition and the corresponding signal preprocessing workflow.

2.1.1. Dataset Description

This study utilized two subsets from the Sleep-EDF Database Expanded to evaluate the performance and generalization ability of the proposed model. Sleep-EDF-20 consists of overnight EEG recordings from 20 healthy adults, with each subject undergoing two consecutive nights of monitoring and approximately eight hours of EEG data collected per night. This subset was used primarily for model training and cross-validation. In contrast,

Sleep-EDF-78 includes single-night or multi-night EEG recordings from 78 different subjects, serving as a benchmark for assessing the cross-subject generalization capability of the model.

All EEG signals were recorded from the Fpz–Cz channel at a sampling rate of 100 Hz. Detailed statistics regarding the number of subjects and the class-wise distribution of sleep stages are presented in Table 1. To ensure robust and reliable evaluation, a 20-fold cross-validation strategy was adopted: in each iteration, one fold was reserved for testing while the remaining folds were used for model training and validation [33].

Table 1. Sleep stage distribution in the datasets.

Dataset	Subjects	Class Samples					Total Samples
		W	N1	N2	N3	REM	
Sleep-EDF-20	20	8285	2804	17,799	5703	7717	42,308
Sleep-EDF-78	78	65,951	21,522	69,132	13,039	25,835	195,479

Figure 1 presents EEG signal samples corresponding to the five sleep stages from the original dataset. Each stage is represented by a single 30 s segment, serving as an individual sample that characterizes the typical EEG activity of that stage. The signals from different stages exhibit distinct variations in both frequency and amplitude, demonstrating clear differences. The Wake (W) stage is characterized by low-amplitude mixed-frequency signals. In contrast, the N1 stage begins to exhibit more pronounced theta waves, accompanied by changes in signal amplitude. The N2 stage is marked by sleep spindles and K-complexes, with the signal showing regular oscillatory patterns. The N3 stage, representing slow-wave sleep, is distinguished by high-amplitude delta waves, while the REM stage displays EEG signals resembling those of the wakefulness stage, with low-amplitude mixed-frequency patterns and rapid eye movements.

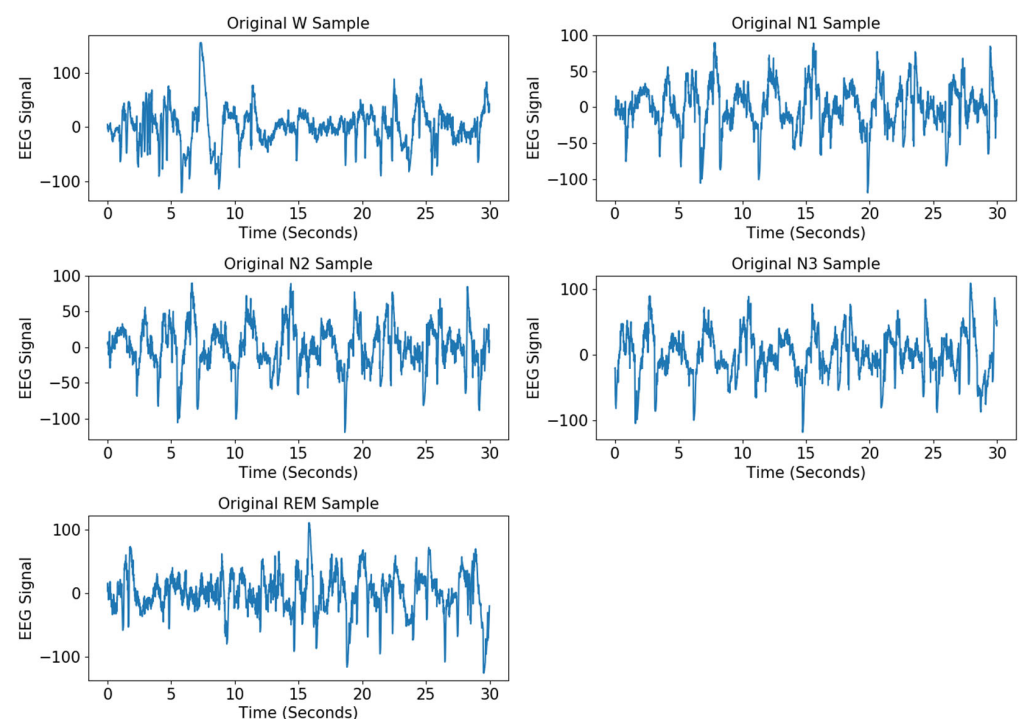


Figure 1. EEG signal examples of the five sleep stages.

To visually assess the separability of traditional time-frequency statistical features across different datasets, we first extracted the average power in four frequency bands: 0.5–4 Hz, 4–8 Hz, 8–13 Hz, and 13–30 Hz for each 30 s segment, along with the mean, standard deviation, and peak-to-peak value of the channel signals, resulting in a total of seven handcrafted features. Subsequently, the feature sets of all samples from the Sleep-EDF-20 and Sleep-EDF-78 subsets were input into the t-SNE algorithm for dimensionality reduction and projection, as shown in Figure 2. It can be observed that the samples from different sleep stages in both datasets are highly mixed in the two-dimensional plane and do not form distinct clusters, indicating that the reliance on the aforementioned time-frequency statistics alone is insufficient for distinguishing the stages in a low-dimensional space. This highlights the necessity of deep models for learning higher-order nonlinear representations.

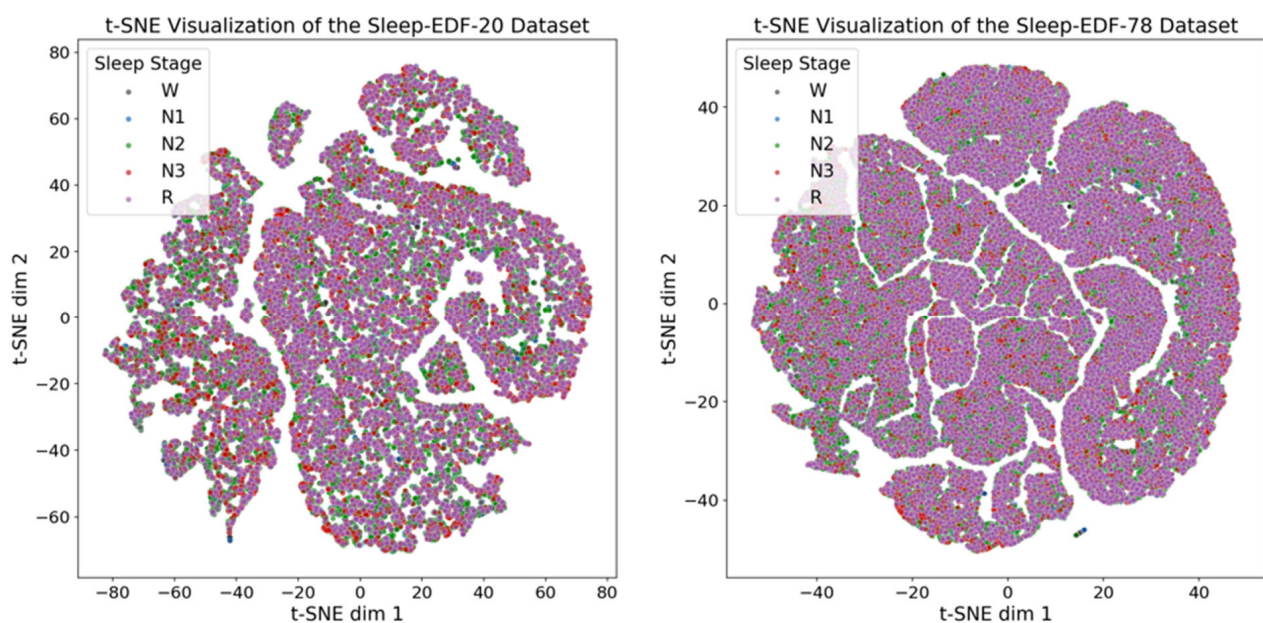


Figure 2. t-SNE analysis of Sleep-EDF-20 and Sleep-EDF-78.

2.1.2. Signal Preprocessing

The preprocessing procedure in this study first extracts EEG signals from specified channels in the raw PSG files and verifies that their sampling rate and recording duration are consistent with the sleep stage information in the annotation files. Then, based on the start time and duration of each segment from the annotation file, the sleep stages (W, N1, N2, N3, REM) and labels such as “MOVE” and “UNK” are mapped to numerical values using a predefined mapping dictionary, and the corresponding labels are assigned to signal segments based on 30 s epochs. After synchronizing the signals with the labels, the first and last non-“W” (wake) epoch indices are located, and 60 epochs (i.e., 30 min) are extended forward and backward to retain the transition data before and after sleep onset. Only the signals and labels within this interval are kept, while long wake segments at both ends are discarded. Subsequently, all epochs labeled as “MOVE” and “UNK” are further removed, resulting in high-quality data that only includes the five standard sleep stages. Finally, the processed signal matrix, label vector, sampling rate, original start time, total duration, and the number of original and filtered epochs, along with other metadata, are saved for subsequent analysis and model training. Figure 3 shows a comparison of the sleep stage distribution before and after preprocessing in a segment of the dataset.

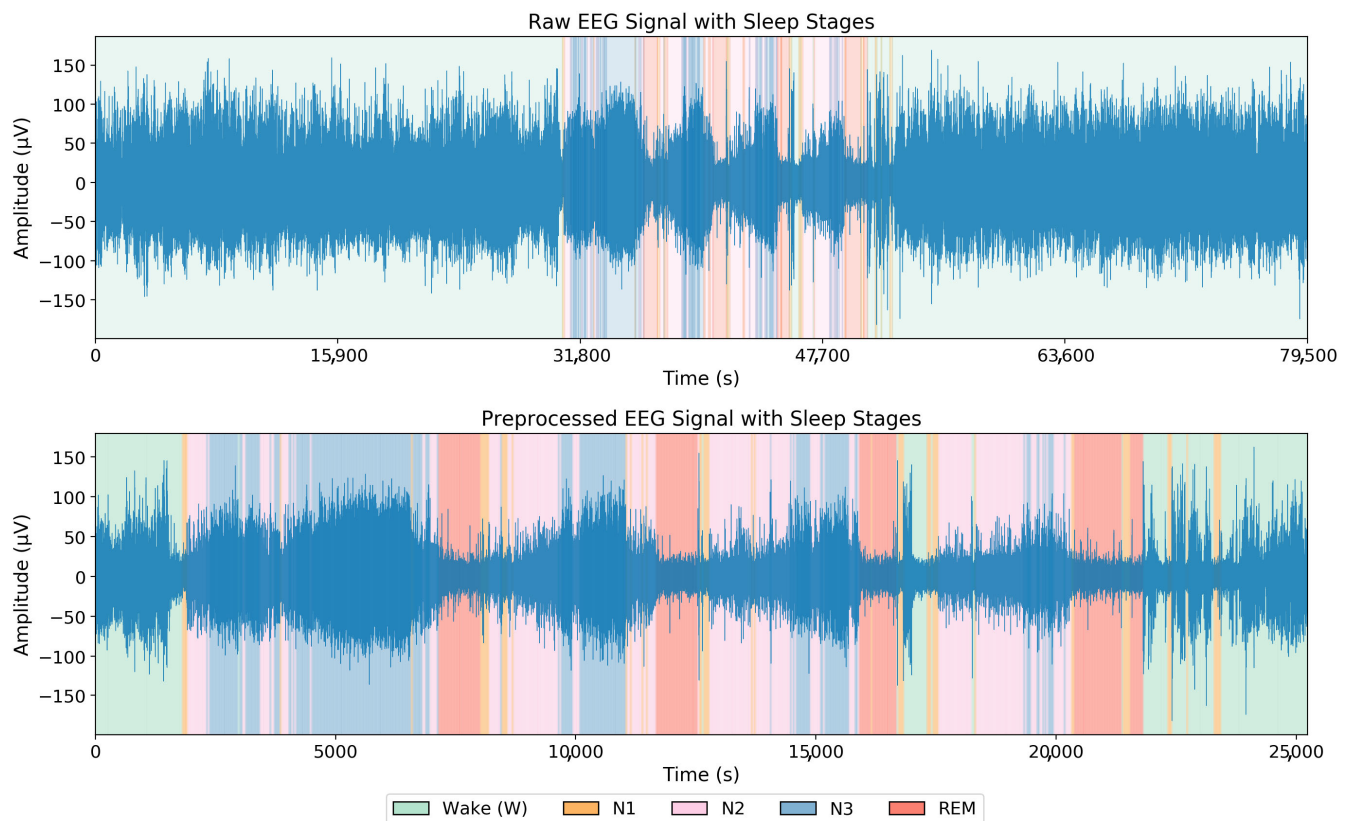


Figure 3. Comparison of Sleep Stage Distributions in Raw Signals and After Preprocessing.

2.2. Methods

ADG-SleepNet is an end-to-end network designed for sleep stage classification, aiming to strike a balance between model performance and lightweight deployment. The overall architecture of ADG-SleepNet is depicted in Figure 4, which illustrates how the network extracts multi-scale temporal features from raw EEG signals through a series of 1D convolutional layers and pooling operations. Subsequently, the innovative Multi-Dilated Gated TCN Module is employed, where the parallel multi-dilation convolutional branches (PMDC Branches) submodule applies multiple depth-wise separable convolutional branches in parallel, each utilizing different dilation rates. This module is further integrated with the adaptive weighted dilation module (AWDM) submodule, which innovatively combines channel attention with an adaptive gating mechanism through an adaptive gating and channel attention mechanism, enabling dynamic selection of the most discriminative channels while adaptively adjusting the weights of each dilation branch based on the input signal, thereby achieving the effective fusion of short-term high-frequency and long-term low-frequency features. Finally, the fused features are processed through a global average pooling layer and a fully connected classification layer to yield the predicted probabilities for the five sleep stages (W, N1, N2, N3, and REM). This end-to-end framework not only achieves accuracy comparable to or surpassing mainstream methods but also significantly optimizes parameter efficiency and computational cost, demonstrating strong potential for practical deployment.

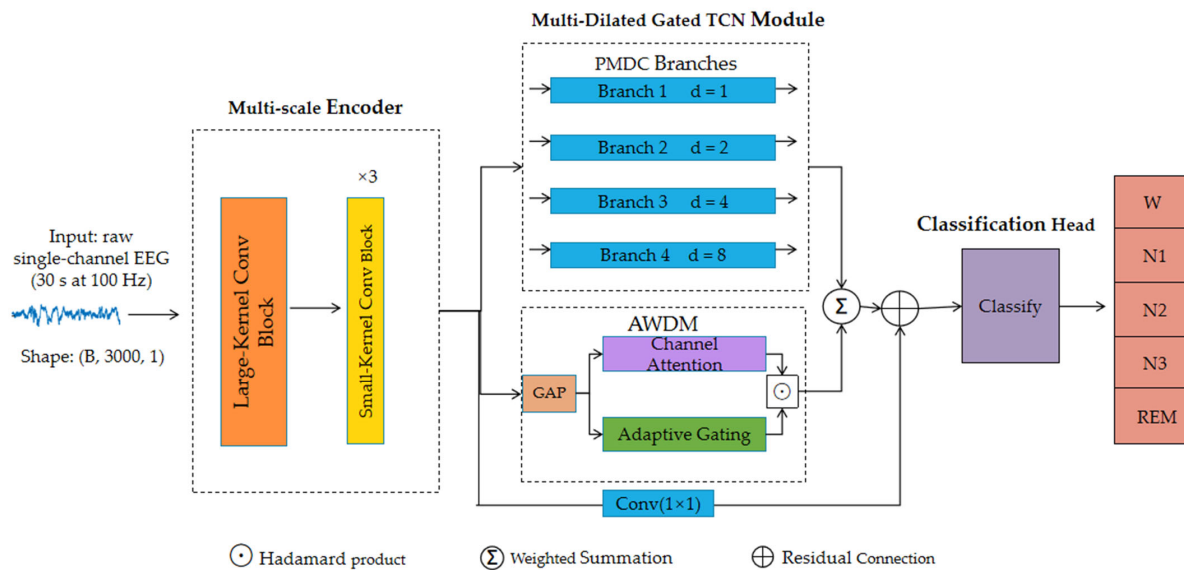


Figure 4. The architecture of the ADG-SleepNet network.

2.2.1. Multi-Scale Encoder

In sleep staging, EEG signals exhibit distinct temporal and spectral characteristics across different sleep stages. For instance, light sleep stages (N1 and N2) are typically characterized by short-duration, high-frequency theta and alpha waves, while deep sleep (N3) is dominated by long-duration, low-frequency delta waves. To effectively capture these varying characteristics, the proposed module employs a two-stage convolutional encoding strategy that integrates multi-scale convolutional designs to efficiently extract temporal features from EEG signals.

As shown in Figure 5, the multi-scale encoder module consists of a two-stage convolutional encoder designed to capture both long-term low-frequency features through large-kernel convolutions and short-term high-frequency features via small-kernel convolutions. The raw EEG signal is first processed by a large-kernel convolutional layer with a kernel size of 50 and a stride of 6. This configuration increases the receptive field, enabling the extraction of coarse-grained, long-term low-frequency features associated with deep sleep stages. The convolution operation is formulated as follows:

$$Y_1 = \text{Conv}(X, W_1, b_1) \quad (1)$$

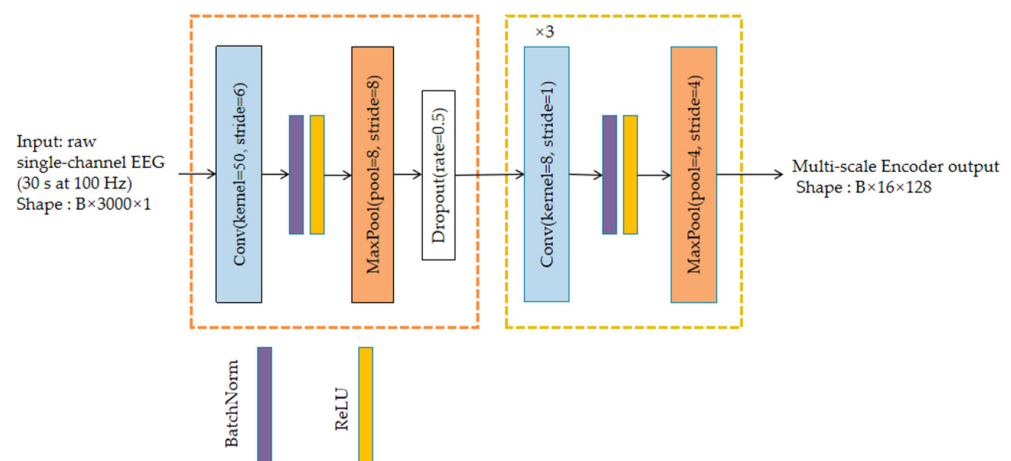


Figure 5. Multi-scale encoder.

Here, X denotes the input signal, W_1 represents the convolutional kernel, and b_1 is the bias term. The output of this operation is subsequently passed through a ReLU activation function:

$$F_1 = \text{ReLU}(\text{BN}(Y_1)) \quad (2)$$

Subsequently, the signal processed by the large-kernel convolution is subjected to a max pooling operation with a window size of 8 and a stride of 8, which further downsamples the signal and smooths out noise. Additionally, a dropout layer with a rate of 0.5 is applied to prevent overfitting:

$$P_1 = \text{MaxPool1D}(F_1, p = 8, s = 8) \quad (3)$$

$$P_1 = \text{Dropout}(P_1, 0.5) \quad (4)$$

Subsequently, the second stage is initiated, in which the input signals undergo further refinement through three consecutive convolutional layers with small kernel sizes (each with a kernel size of 8 and a stride of 1). The design of these three small-kernel convolutional layers enables the effective capture of rapidly changing high-frequency details in EEG signals, particularly the wave features characteristic of light sleep stages. The small-kernel convolution operations are defined as follows:

$$Y_2 = \text{Conv}(P_1, W_1, b_2) \quad (5)$$

$$Y_3 = \text{Conv}(Y_2, W_3, b_2) \quad (6)$$

$$Y_4 = \text{Conv}(Y_3, W_4, b_4) \quad (7)$$

After each small-kernel convolutional layer, the output is processed by a ReLU activation function:

$$F_i = \text{ReLU}(\text{BN}(Y_i)) \quad i = 2, 3, 4 \quad (8)$$

The signals following the small-kernel convolutions are subsequently subjected to a max pooling operation (with a pooling window size of 4 and a stride of 4) for dimensionality reduction, which further extracts high-frequency features and enhances the model's generalization capability:

$$P_2 = \text{MaxPool1D}(F_4, p = 4, s = 4) \quad (9)$$

At the final stage of this module, the output features are prepared for the subsequent adaptive dilation-rate gated TCN module. In this manner, the multi-scale encoder demonstrates exceptional capability in extracting features at multiple scales, thereby providing richer representations for the downstream network layers and significantly enhancing the model's classification performance. The architecture illustrated in Figure 5 clearly delineates this multi-scale convolutional process, enabling the network to effectively capture both low-frequency (deep sleep) and high-frequency (light sleep) signal characteristics, thus facilitating accurate classification of complex sleep stages.

2.2.2. PMDC Branches

In temporal data modeling, capturing dependencies across multiple time scales is essential. To tackle this, we propose a method using multi-dilated convolutional branches, where multiple dilated convolutions are applied in parallel to extract features at different scales. This approach enhances the network's ability to capture temporal dependencies across various time spans.

Traditional convolutional neural networks (CNNs) rely on a fixed receptive field, limiting their ability to model long-range temporal dependencies. While dilated convolutions help expand the receptive field without increasing computational cost or stride, using a single dilation rate may still be insufficient for complex temporal data. To address this, we introduce multi-dilated convolutional branches, where each branch uses a different dilation rate, allowing the model to capture features at multiple temporal scales simultaneously. This enables the network to effectively perceive dependencies over a wide range of time scales.

Nevertheless, using a single dilation rate may still be inadequate when dealing with complex temporal data. To simultaneously capture information across various time scales, we design multi-dilated convolutional branches, in which multiple convolutional branches, each with a different dilation rate, are employed in parallel. Each branch is able to extract temporal features at a specific scale, thereby equipping the model with strong perceptual capabilities across a broad range of time scales.

In sleep EEG signals, short-term high-frequency oscillations (lasting several hundred milliseconds) often coexist with cross-epoch low-frequency δ waves (lasting from several seconds to over ten seconds). Convolutions with a single receptive field struggle to capture both simultaneously. To this end, as illustrated in Figure 6, we design four parallel depth-wise separable dilated convolutional branches with dilation rates of $d \in \{1, 2, 4, 8\}$, thereby covering multi-scale temporal dependencies.

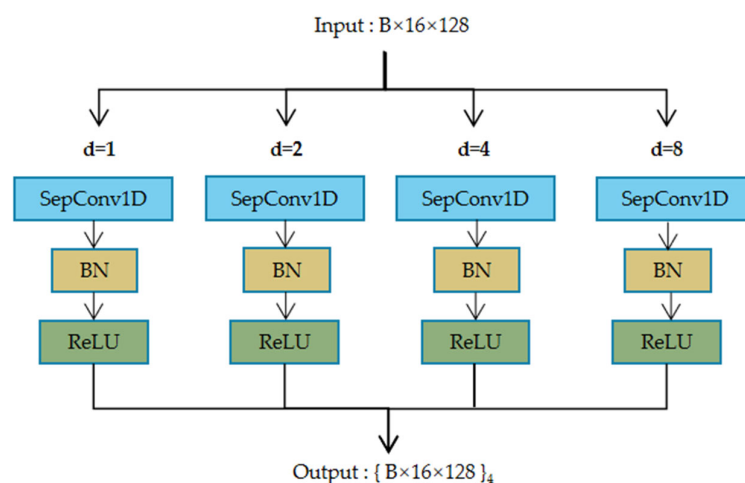


Figure 6. Architecture of the PMDC Branches.

Let the feature map output from the previous stage be denoted as:

$$h_{\text{seq}} \in \mathbb{R}^{l \times c} \quad (10)$$

where l denotes the temporal length and c represents the number of channels. The output of the i -th branch is denoted as:

$$Y_i = \text{SeparableCpvn1D}(h_{\text{seq}}; f = 128, k = 3, d_i) \quad (11)$$

Here, the number of output channels is set to $f = 128$ to ensure compatibility with the subsequent classifier; the convolutional kernel length is set to $k = 3$, which strikes a balance between local feature extraction and computational efficiency; and the dilation rate d_i determines the receptive field:

$$\text{RF}_i = (k - 1) \times d_i + 1 \quad (12)$$

Accordingly, the receptive fields of the four branches are 3, 5, 9, and 17 time steps, respectively, allowing for the parallel capture of key signal fluctuations from the millisecond to the second scale. The rationale for employing depth-wise separable convolutions lies in their ability to decouple spatial and pointwise convolutions, thereby reducing the parameter count and computational cost of standard convolutions to the order of $f \times k \times c / (k \times c + f)$. This design significantly decreases the model size and accelerates inference speed, while preserving the advantages of the multi-branch structure. Ultimately, the outputs from each branch are fused with the outputs of the subsequently described adaptive weighted dilation-rate module.

2.2.3. AWDM

Capturing dependencies across different time scales is crucial for improving the model's expressive power and accuracy in temporal data modeling. Conventional convolutional neural networks (CNNs) typically rely on fixed receptive fields, which are effective for capturing local information but limited when processing temporal data with multi-scale dependencies. To address this, we propose the adaptive weighted dilation-rate module. Building on multi-dilated convolutional parallel branches, this module integrates an adaptive gating mechanism with channel attention to dynamically assign weights to each dilation-rate branch, enhancing the model's ability to capture features at multiple temporal scales.

The core idea of this module is to first extract multi-scale features using dilated convolutions from the parallel branches, and then adjust the weights of each branch through the adaptive gating mechanism. This enables the network to dynamically prioritize branches based on the input signal's global characteristics. Additionally, the channel attention mechanism refines the focus on key features, selectively strengthening the contribution of important channels within multi-scale temporal representations. This approach allows the model to effectively adapt to both short-term fluctuations and long-term dependencies, capturing a broad range of temporal information. In the adaptive dilation rate gating mechanism, the input signal is first processed through global average pooling to obtain a global feature vector for each sample. This global feature vector is then passed through a fully connected layer, which calculates the gating weights for each branch. These weights reflect the importance of each branch for the current task. Specifically, the weight coefficients learned by this fully connected layer are dynamically adjusted based on the characteristics of the input signal, thereby automatically assigning appropriate weighted coefficients to each dilation rate branch.

The AWDM structure is shown in Figure 7. The output feature tensor of the encoder is $X \in \mathbb{R}^{B \times T \times C}$, where B is the batch size, T is the number of time steps, and C is the number of channels. AWDM first performs global average pooling over the time and channel dimensions for each sample, resulting in the compressed sample-level channel description vector:

$$p_b = \frac{1}{T \times C} \sum_{t=1}^T \sum_{c=1}^C X_{b,t,c}, b = 1, 2, \dots, B \quad (13)$$

The pooled feature p_b represents the overall activation level of each sample, serving as the foundation for attention allocation.

As shown in Figure 7, the AWDM adopts a parallel dual-branch structure: the left side is the channel attention path, and the right side is the adaptive gating path. Both paths take the global pooling result p_b as input and generate dilation rate weights through independent fully connected networks.

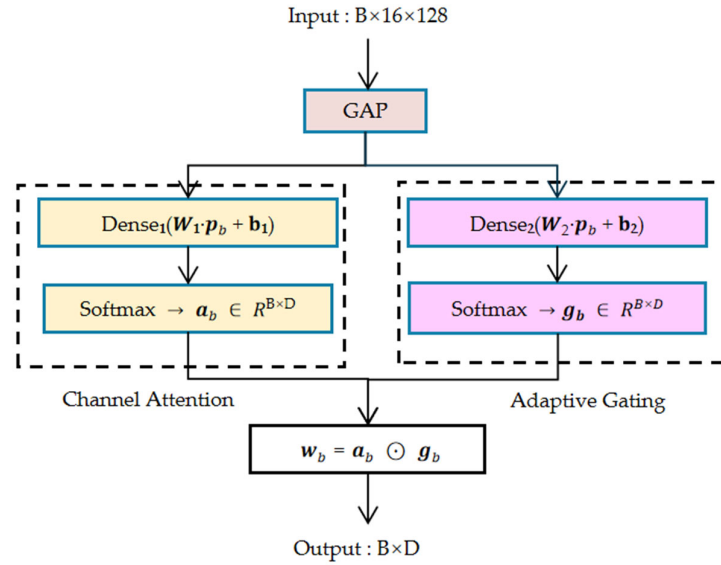


Figure 7. Architecture of the adaptive weighted dilation-rate module (AWDM). Here, b_1 and b_2 are the learnable bias parameters in the attention and gating sub-networks, respectively.

The attention vector is obtained by applying a dense layer and a Softmax activation to the global pooled descriptor:

$$a_b = \text{Softmax}(\text{Attention}(p_b)) \in \mathbb{R}^{B \times D} \quad (14)$$

Here, D is the number of dilation branches. The gating branch employs an independent set of weights to compute the control distribution via Softmax:

$$g_b = \text{Softmax}(\text{Dense}(p_b)) \in \mathbb{R}^{B \times D} \quad (15)$$

This mechanism further enhances the network's ability to focus on critical temporal features by emphasizing those channels that are most important for the task. For example, in certain scenarios, the features of some channels may contain richer high-frequency fluctuation information, whereas others may capture low-frequency dependencies. The channel attention mechanism enables the network to adaptively assign appropriate weights to these different channels, ensuring that the contribution of important channels is reinforced when capturing dependencies at different temporal scales.

By combining the gating weights G and the attention weights A , the final weighting coefficient $w \in \mathbb{R}^{B \times D}$ for each dilation-rate branch is obtained, which is calculated as follows:

$$w_b = g_b \odot a_b \quad (16)$$

The resulting features are subsequently fed into the underlying Weighted Summation module for further computation.

2.2.4. Weighted Summation

In the previous section, we introduced the adaptive weighted dilation-rate module (AWDM), which produces attention and gating vectors through two parallel branches. These vectors are fused via an element-wise Hadamard product to generate the final weighting coefficients, which modulate the contribution of each dilation-rate branch during the feature fusion process.

This section details how the outputs from different dilation-rate branches are integrated using the learned weights to form a unified multi-scale feature representation.

Let the output of the d -th dilated-convolution branch be denoted as $Y^{(d)} \in \mathbb{R}^{B \times T \times C}$, where B , T , and C represent the batch size, temporal length, and number of channels, respectively. To enable dimension alignment for branch-wise weighting, we stack the outputs of all branches along an additional branch axis and reshape the sample-wise weights produced by AWDM to have singleton time and channel dimensions, so that they broadcast over these axes and align with the features. This yields element-wise compatible shapes and allows direct Hadamard multiplication. Consequently, for the b -th sample, the weighted output of branch d is given by:

$$\widehat{Y}_b^{(d)} = w_{b,d} \cdot Y_b^{(d)} \quad (17)$$

The final fused representation is obtained by summing the weighted branch outputs along the branch axis, thereby reducing the stacked tensor back to $B \times T \times C$; this operation preserves the temporal and channel resolutions and performs only sample-wise adaptive reweighting of the contributions from different dilation rates:

$$F_b = \sum_{d=1}^D \widehat{Y}_b^{(d)} \quad (18)$$

This adaptive fusion mechanism enables the model to dynamically emphasize different temporal receptive fields based on the input, thereby capturing both short-term and long-term temporal dependencies more effectively.

To further enhance feature representation and gradient flow, a residual connection is introduced, as illustrated in Figure 4. Specifically, the encoder output X_b is first projected to match the dimensionality of F_b using a one-dimensional convolution:

$$X_b^{\text{proj}} = \text{Conv1D}(X_b) \quad (19)$$

The residual-enhanced feature is then obtained via element-wise addition:

$$F_b^{\text{res}} = F_b + X_b^{\text{proj}} \quad (20)$$

This operation preserves the original structural information while enabling deep integration with the multi-scale temporal features, thus improving both the expressive power and training stability of the model. Finally, the fused feature F_b^{res} is passed through a ReLU activation and fed into the downstream classifier for final prediction. The seamless integration of adaptive fusion and residual connection allows the model to effectively aggregate multi-scale temporal information and adapt to various types of temporal dependencies, ultimately enhancing classification performance and generalization capability.

2.2.5. Classification Head

In this model, the input to the Classification Head module is first subjected to global average pooling along the temporal dimension. The purpose of this operation is to compress the temporal dimension and retain the global information of each feature channel, thereby reducing the computational complexity of subsequent layers and enhancing the expressive power of the features.

Subsequently, the globally pooled features are passed through a fully connected layer, whose output dimension is set to 5, corresponding to the five sleep stages (W, N1, N2, N3, REM). This fully connected layer is immediately followed by a Softmax activation function, which transforms the predicted scores (logits) for each class into a probability distribution. The Softmax function ensures that the predicted probability for

each class lies within the range $[0, 1]$, and that the sum of the probabilities across all classes equals 1.

In this way, the model outputs the predicted probability for each sleep stage. Let the output be $p = [p_W, p_{N1}, p_{N2}, p_{N3}, p_{REM}]$, where each p_i denotes the predicted probability for the corresponding sleep stage. The final predicted sleep stage is determined by selecting the class with the highest probability.

Through this process, ADG-SleepNet achieves end-to-end multi-scale dynamic fusion and sleep stage prediction. This framework integrates multi-level temporal feature extraction and adaptive weighting mechanisms, enabling the model to effectively recognize and classify different sleep stages, thereby providing precise support for sleep research and health monitoring.

Algorithm 1 shows the pseudocode of the ADG-SleepNet network architecture.

Algorithm 1 Pseudocode for ADG-SleepNet

Input: Time steps $T0 = 3000$, Batch size B , Channels $C0 = 1$, Classes $K = 5$, Branches $D = 4$;
Output: Trained/Inference model;
1: function BUILD_ADG_SLEEPNET($T0, B, C0, K, D$)
2: $X \in \mathbb{R}^{B \times T0 \times C0}$;
3: $ip \leftarrow \text{Input}(\text{shape} = (T0, C0))$;
Encoder Block:
4: $H1 \leftarrow \text{Conv1D}(k = 50, s = 6)(ip) \rightarrow \text{BN} \rightarrow \text{ReLU}$;
5: $H1 \leftarrow \text{MaxPool1D}(\text{win} = 8, s = 8) \rightarrow \text{Dropout}(0.5) (H1)$;
6: $H1 \leftarrow 3 \times [\text{Conv1D}(k = 8, s = 1) \rightarrow \text{BN} \rightarrow \text{ReLU}] (H1)$;
7: $X_{\text{enc}} \leftarrow \text{MaxPool1D}(\text{win} = 4, s = 4)(H1)$;
PMDC Block (parallel):
8: $\text{dilations} \leftarrow [1, 2, 4, 8]$;
9: for d in dilations do
10: $Z[d] \leftarrow \text{DWSepConv1D}(k = 3, \text{dilation} = d, \text{out} = 128) (X_{\text{enc}}) \rightarrow \text{BN} \rightarrow \text{ReLU}$;
11: end for
AWDM Block (Channel Attention(parallel to)Adaptive Gating):
12: $g \leftarrow \text{GAP}_t (X_{\text{enc}})$;
13: $\omega \leftarrow \text{ChannelAttention}(g) \odot \text{AdaptiveGating}(g)$;
Weighted Summation:
14: $Z_{\text{stack}} \leftarrow \text{stack} (\{Z[d]\}_{d \in \text{dilations}}, \text{axis} = \text{branches})$;
15: $\omega^\# \leftarrow \text{expand} (\omega \rightarrow B \times 1 \times 1 \times D)$;
16: $Z^\sim \leftarrow \sum_{d \in \text{dilations}} (Z_{\text{stack}}[:, :, :, d] * \omega^\#[:, :, :, d])$;
Residual and Activation:
17: $Y \leftarrow \text{ReLU}(Z^\sim + \text{Conv1} \times 1(X_{\text{enc}}))$;
Output:
18: $h \leftarrow \text{GAP}_t(Y)$;
19: $\text{probs} \leftarrow \text{Softmax}(\text{Dense}(K)(h))$;
20: $\text{model} \leftarrow \text{Model}(ip, \text{probs})$;
21: return model;
22: end function

2.3. Experimental Setup

To systematically evaluate the effectiveness and robustness of ADG-SleepNet in the context of automatic sleep staging, this study strictly adheres to mainstream experimental protocols for single-channel EEG-based classification, with careful consideration of data segmentation, training procedures, model hyperparameters, and optimization strategies.

Specifically, experiments are conducted on both the Sleep-EDF-20 and Sleep-EDF-78 subsets, employing a unified 20-fold cross-validation scheme [35]. In each experimental round, one of the 20 folds is used as an independent test set, while the remaining 19 folds serve for model training and validation, thereby maximizing the assessment of model generalization and result robustness. To further mitigate overfitting, 10% of the training set is randomly selected as a validation set, enabling real-time monitoring of model performance on unseen data and facilitating adaptive early stopping based on validation loss dynamics [36], which enhances both training efficiency and generalization capability.

During the model training process, the raw EEG signals are first segmented into 30 s temporal epochs (each containing 3000 samples) and fed into the multi-scale encoder module to extract local feature sequences. These feature sequences are then input into subsequent modules for modeling.

To enhance the model's generalization ability, lightweight data augmentation strategies, such as segment permutation and time reversal, are applied during the training phase. The Adam optimizer is used with an initial learning rate of 1×10^{-4} and a decay factor of 1×10^{-8} . The learning rate is dynamically adjusted based on the validation set performance during training. To prevent gradient explosion, all gradients are clipped with a maximum norm of 5.0. Additionally, L2 regularization is applied to the convolutional layer weights with a weight decay coefficient of 1×10^{-3} to further improve the model's generalization.

Training is conducted for a maximum of 150 epochs with an early stopping window set to 30 epochs: if there is no significant improvement in the validation set metrics within 30 consecutive epochs, training is automatically terminated. All experiments are conducted with a mini-batch size of 15, balancing training efficiency and model performance. Throughout the training and testing processes, multiple metrics, including overall accuracy, macro-average F1 score, per-class F1 scores, and model parameters, are continuously monitored and recorded for final model evaluation and comparison. All experimental procedures and hyperparameter settings are kept consistent across different datasets, without any manual intervention, ensuring fairness and reproducibility of the evaluation.

2.4. Evaluation Metrics

To comprehensively and objectively evaluate the practical performance of the proposed model on sleep stage classification tasks, we designed and adopted a suite of evaluation metrics addressing overall classification accuracy, per-stage discriminative capability, and model architectural complexity. By conducting a detailed assessment across multiple dimensions, we aim to thoroughly capture the model's real-world applicability and its comparative advantages and limitations relative to existing methods. The specific metrics and their corresponding formulations are as follows:

First, Accuracy (Acc) serves as the most straightforward indicator of model performance, reflecting the overall proportion of correctly classified samples. It is defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

In this context, TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives. Acc provides an intuitive measure of the model's performance on the overall task; however, in cases of class imbalance, Acc may be influenced by the dominance of certain classes, thus necessitating a more comprehensive analysis through the incorporation of additional metrics.

Furthermore, precision (Pre) and recall (Rec) evaluate the model's performance in terms of its accuracy when predicting positive cases and the proportion of actual positive samples correctly identified by the model, respectively. The formula for Pre is as follows:

$$Pre = \frac{TP}{TP + FP} \quad (22)$$

The formula for Rec is as follows:

$$Rec = \frac{TP}{TP + FN} \quad (23)$$

Pre emphasizes the accurate exclusion of negative samples, while Rec focuses on the model's ability to identify as many positive samples as possible. In many practical scenarios, these two metrics often exhibit a trade-off, where improving one may lead to a reduction in the other. Therefore, a balanced approach is required for their effective use.

To comprehensively consider both Pre and Rec, we adopt the F1 Score (F1), which is the harmonic mean of these two metrics. The formula is as follows:

$$F1 = 2 \times \frac{Pre \times Rec}{Pre + Rec} \quad (24)$$

The F1 provides a reasonable balance between Precision and Recall, making the evaluation results more comprehensive.

In addition, the Macro F1 Score (MF1) effectively addresses class imbalance by calculating the F1 for each class and taking the arithmetic mean of all class scores. The formula is as follows:

$$MF1 = \frac{1}{N} \sum_{i=1}^N F1_i \quad (25)$$

Here, N represents the number of classes, and $F1_i$ is the F1 for the i-th class. The MF1 gives equal weight to the performance of each class, ensuring that the model's performance across all classes is evaluated fairly.

By using these metrics in combination, we are able to assess the model's performance comprehensively from multiple perspectives and gain deeper insights into its behavior across scenarios. These results not only help identify the strengths and weaknesses of the model but also provide guidance for future optimization.

3. Results

This chapter begins by presenting the overall performance of ADG-SleepNet on both the Sleep-EDF-20 and Sleep-EDF-78 subsets. Next, the proposed method is comprehensively benchmarked against several recent state-of-the-art approaches to highlight its relative advantages. To further elucidate the effectiveness of each component, ablation experiments are conducted. Lastly, confusion matrix analyses are employed to provide an in-depth examination of the model's strengths and limitations across different sleep stages.

3.1. Model Performance

To systematically evaluate the overall performance of the proposed ADG-SleepNet model on automatic sleep staging, rigorous experiments were conducted on two public datasets: Sleep-EDF-20 and Sleep-EDF-78. All experiments employed 20-fold cross-validation to ensure the robustness and generalizability of the results.

Specifically, on the Sleep-EDF-20 dataset, ADG-SleepNet achieved an average Acc of 87.1% and MF1 score of 84.0%. These results indicate that the model is capable of

accurately distinguishing between different sleep stages, reaching a state-of-the-art level of classification performance among comparable approaches.

Furthermore, on the larger and more complex Sleep-EDF-78 dataset, ADG-SleepNet consistently demonstrated strong generalization capabilities, achieving an average Acc of 85.1% and MF1 of 81.1%. These findings further illustrate that ADG-SleepNet can not only accommodate inter-subject physiological variability, but also effectively handle distributional shifts and challenges posed by large-scale data.

ADG-SleepNet achieves 0.011 G FLOPs and 0.49 M parameters on the NVIDIA V100 (32 GB) GPU, with an inference latency of 1.84 ms. These results demonstrate that the model not only maintains high computational efficiency but also achieves significant lightweight design while preserving accuracy.

3.2. Comparison with Other Methods

To comprehensively evaluate the effectiveness of the ADG-SleepNet model in automatic sleep staging, we conducted comparisons with various state-of-the-art and recent lightweight deep learning models on the Sleep-EDF-20 and Sleep-EDF-78 datasets, as shown in Table 2. The benchmark models include representative methods in the sleep staging domain, such as SleepEEGNet, DeepSleepNet-lite, IITNet, AttnSleep, L-SeqSleepNet, SeriesSleepNet, TinySleepNet, SleepTransformer, EEGSNet, FFTCCN, as well as recently proposed models like MultiSEss, ZleepAnlystNet, and S4Sleep.

Table 2. Comparison of Model Performance.

Datasets	Methods	Acc	MF1	Per-Class F1-Score (F1)					Parameters ×10 ⁶
				W	N1	N2	N3	REM	
Sleep-EDF-20	SleepEEGNet [37]	84.3	79.7	89.2	52.2	86.8	85.1	85.0	2.60
	DeepSleepNet-lite [38]	84.0	78.0	87.1	44.4	87.9	88.2	82.4	0.60
	IITNet [39]	83.9	77.6	87.7	43.4	87.7	86.7	82.5	
	AttnSleep [29]	85.6	80.9	90.3	47.9	89.8	89.0	85.0	4.54
	L-SeqSleepNet [40]	86.3	79.3	91.6	45.3	88.5	86.2	85.2	0.63
	SeriesSleepNet [41]	84.8	79.8	89.7	49.4	87.9	88.2	83.6	
	MultiSEss [42]	83.8	73.4	89.5	45.1	87.6	79.0	78.0	
	ZleepAnlystNet [43]	84.1	78.3	93.1	48.3	85.9	81.2	82.9	
	S4Sleep [44]	84.4	80.7	92.3	55.8	86.1	82.1	87.7	
	ADG-SleepNet	87.1	84.0	93.5	58.4	92.1	88.8	87.2	0.49
Sleep-EDF-78	SleepEEGNet	80.0	73.6	91.7	44.1	82.5	73.5	76.1	2.60
	TinySleepNet [45]	83.1	78.1	92.8	51.0	85.3	81.1	80.3	1.3
	DeepSleepNet-lite	80.3	75.2	91.5	46.0	82.9	79.2	76.4	0.60
	AttnSleep	82.9	78.1	92.6	47.4	85.5	83.7	81.5	4.54
	SleepTransformer [46]	81.4	74.3	91.7	40.4	84.3	77.9	77.2	3.7
	EEGSNet [47]	83.0	77.3	93.2	50.0	84.2	74.4	83.5	0.6
	MultiSEss	82.3	72.2	91.5	47.6	83.5	75.3	75.1	
	FFTCN [48]	82.6	77.1	92.2	47.3	84.8	80.0	81.0	
	ADG-SleepNet	85.1	81.1	91.3	53.7	90.0	85.5	84.8	0.49

In terms of overall performance, ADG-SleepNet attained an Acc of 87.1% and an MF1 of 84.0% on the Sleep-EDF-20 dataset, surpassing most mainstream models by a margin of 1–3 percentage points. These results highlight the model's superior consistency and reliability in classifying different sleep stages. Compared to established models such as AttnSleep and SleepEEGNet, ADG-SleepNet consistently achieved leading or near-leading results in both Acc and MF1. Specifically, AttnSleep reported an Acc of 85.6% and an MF1 of 80.9%, while SleepEEGNet achieved 84.3% Acc and 79.7% MF1. Other lightweight competitors, including DeepSleepNet-lite, IITNet, and L-SeqSleepNet, demonstrated comparatively lower overall performance. In comparison to more recent models, ADG-SleepNet also

outperformed MultiSEss, which achieved an Acc of 83.8% and an MF1 of 79.0%, Zleep-AnlystNet, with an Acc of 84.1% and an MF1 of 82.9%, and S4Sleep, which attained an Acc of 84.4% and an MF1 of 87.7%. These findings collectively indicate that ADG-SleepNet delivers breakthrough results without depending on large or complex model architectures.

Remarkably, for the most challenging N1 stage in the Sleep-EDF-20 dataset, ADG-SleepNet achieved an F1 score of 58.4%, considerably outperforming leading mainstream methods. By comparison, the F1 scores of representative models such as AttnSleep, SleepEEGNet, and L-SeqSleepNet for the N1 stage typically ranged from 44% to 53%. This clearly underscores the strong capacity of ADG-SleepNet to handle highly imbalanced samples and complex feature entanglement in sleep staging. Furthermore, the model's overall Acc and MF1 on the Sleep-EDF-20 test set (87.1% and 84.0%, respectively) not only reinforce its effectiveness in terms of improving recognition of the N1 stage but also demonstrate robust generalization for the global sleep staging task.

Nevertheless, certain limitations persist in recognizing the N1 stage. The F1 score for N1 remains notably lower than those for W and N2, suggesting that misclassification still occurs under severe class imbalance and overlapping physiological features. Moreover, the model's ability to delineate transitional states (e.g., the N1–N2 boundary) requires further enhancement. Future research will focus on improving the representation of rare stages and incorporating more multimodal and temporal contextual information to bolster robustness and generalization in complex real-world scenarios.

It is also worth emphasizing that ADG-SleepNet achieves outstanding model compression. With only 0.49 million parameters, it ranks among the most compact models in its class. For comparison, SleepEEGNet and AttnSleep comprise 2.60 M and 4.54 M parameters, respectively, while even other lightweight models such as DeepSleepNet-lite and TinySleepNet exceed 0.6 M parameters. This significant reduction demonstrates the lightweight architectural advantage of ADG-SleepNet, substantially lowering storage and computational costs without compromising staging accuracy, thereby making it highly suitable for deployment on mobile and embedded platforms.

In summary, ADG-SleepNet not only consistently surpasses existing models in overall and stage-wise accuracy and F1 scores but also offers advantages in model complexity and practical deployment. In particular, its superior performance in the challenging N1 stage provides strong support for clinical applications in sleep medicine and portable monitoring devices. Compared with current mainstream lightweight models, ADG-SleepNet achieves a better balance between performance and efficiency, fully showcasing the advancements and practical significance of the proposed method.

3.3. Ablation Study

To further assess the contribution of each key component in ADG-SleepNet to the overall performance, we conducted ablation experiments by progressively removing the adaptive gating (AG) and channel attention (CA) modules, and evaluated their impact on the Sleep-EDF-20 dataset. The detailed results are presented in Table 3.

Table 3. Ablation Study Comparison Table on the Sleep-EDF-20 Dataset.

Configuration	Acc	MF1	Per-Class F1-Score (F1)				
			W	N1	N2	N3	REM
ADG-SleepNet	87.1	84.0	93.5	58.4	92.1	88.8	87.2
–AG	85.3	79.1	92.1	47.6	89.4	86.1	80.2
–CA	86.0	79.8	93.1	49.5	90.3	88.0	77.9
–(AG+CA)	84.1	78.8	91.3	44.2	88.7	85.3	75.2

As shown by the experimental results, the complete model—incorporating both the AG and CA modules—achieved the highest performance, with an Acc of 87.1% and a macro-averaged MF1 of 84.0%. When the adaptive gating module (−AG) was removed, the model’s performance degraded significantly, with Acc and MF1 dropping to 85.3% and 79.1%, respectively, highlighting the crucial role of adaptive gating in feature selection and information flow. Further, removing only the channel attention module (−CA) resulted in Acc and MF1 values of 86.0% and 79.8%, respectively—also lower than those of the full model, but higher than those observed when the AG module was removed. As shown in Figure 8, the average attention weights of the top 3 channels exhibit a continuous increase during the training process, gradually stabilizing over time. This indicates that the channel attention mechanism can dynamically focus on key channels, significantly enhancing their feature contributions. These results suggest that the channel attention module effectively models inter-channel feature dependencies, thereby improving the overall performance of the model. Notably, this enhancement is evident in both Acc and macro-average MF1, with a particularly marked improvement in MF1.

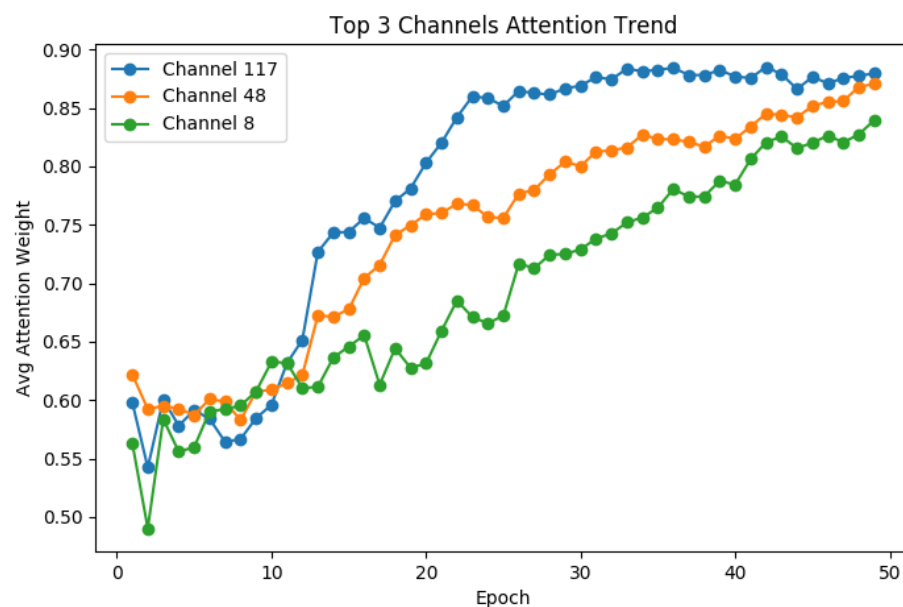


Figure 8. Top 3 attention weight trends.

Moreover, when both the AG and CA modules were simultaneously removed (AG+CA), the model’s performance further declined, with Acc reduced to 84.1% and MF1 to 78.8%. These results not only demonstrate the complementary relationship between the two modules but also confirm that the superior performance of ADG-SleepNet stems from the collaborative enhancement of feature representation provided by multiple mechanisms.

3.4. Confusion Matrix Analysis

To gain a comprehensive understanding of the discriminative capabilities of ADG-SleepNet across different sleep stages, as well as its specific misclassification tendencies, Figure 9 displays the confusion matrix for the model on the Sleep-EDF-20 test set. The results indicate that ADG-SleepNet achieves notably high recognition rates along the main diagonal, underscoring its effectiveness in correctly classifying the majority of samples across various sleep stages. However, the distribution of off-diagonal elements highlights the inherent challenges in sleep staging and offers valuable insights into the detailed classification behaviors of the model.

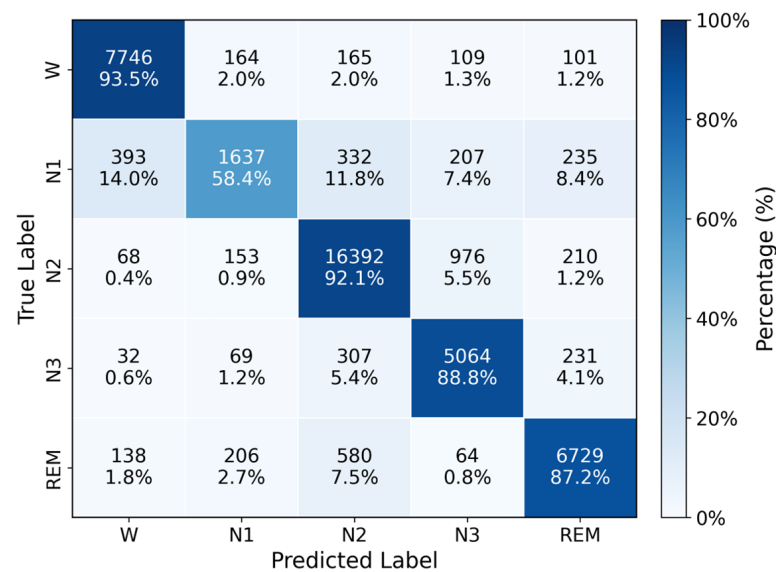


Figure 9. Confusion Matrix of ADG-SleepNet on the Sleep-EDF-20 Dataset.

In particular, the model demonstrates exceptional performance in identifying the Wake (W), N2, and REM stages, with main diagonal accuracies reaching 93.5%, 92.1%, and 87.2%, respectively. These results reflect the model's robust ability to accurately detect wakefulness, typical non-REM, and REM sleep periods. For the N3 stage, the main diagonal accuracy is slightly lower at 88.8%, but still remains at a competitive level, indicating strong recognition of deep sleep patterns. By contrast, the performance for the N1 stage is more modest, with a main diagonal accuracy of 58.4%—the lowest among all stages. This observation is consistent with established clinical findings, as N1 is both underrepresented and characterized by highly overlapping features with adjacent stages, thus presenting a persistent challenge in automated sleep staging.

Further analysis of the confusion matrix reveals that misclassifications of N1 predominantly occur as W, N2, and REM, with respective rates of 14.0%, 11.8%, and 8.4%. This pattern underscores the difficulties in discriminating transitional states, particularly between N1 and its neighboring stages, largely due to the ambiguous physiological nature of N1 and the considerable overlap in EEG characteristics. Additionally, there is a notable degree of mutual misclassification between N2 and N3 (5.5% of N2 predicted as N3, and 5.4% of N3 predicted as N2), suggesting that further refinement is needed to better distinguish between light and deep NREM sleep stages. It is also observed that a proportion of REM samples are misclassified as N2 (7.5%), highlighting the partial feature overlap that exists between certain sleep stages within complex sleep architectures.

In summary, ADG-SleepNet achieves high classification accuracy across the major sleep stages, thereby promoting overall staging consistency and reliability. Importantly, despite the substantial challenges associated with N1, the model still demonstrates commendable recognition capability in this stage. Compared to existing state-of-the-art baselines, the confusion matrix of ADG-SleepNet exhibits a more balanced distribution, with fewer extreme misclassification cases, further supporting its stability and practical application potential. Nevertheless, the discrimination between N1 and its neighboring stages remains an area for improvement. Future research will aim to address these challenges by incorporating multimodal feature integration and exploring finer-grained temporal context modeling, thereby further enhancing recognition accuracy for complex transitional states.

4. Discussion

The proposed ADG-SleepNet demonstrates significant performance improvements in single-channel EEG-based sleep staging, notably achieving higher staging accuracy and macro-averaged F1 scores than mainstream lightweight models, even with an extremely compact parameter count. Based on the aforementioned experimental results, the model's performance can be further discussed from several perspectives.

Firstly, ADG-SleepNet achieves outstanding overall performance on both the Sleep-EDF-20 and Sleep-EDF-78 benchmark datasets. Specifically, the model attains an accuracy of 87.1% and a macro-averaged F1 score of 84.0% on Sleep-EDF-20, and 85.1% accuracy with 81.1% macro-F1 on Sleep-EDF-78. These results not only surpass most lightweight neural network models but also demonstrate notable advantages in challenging stages such as N1. Analyses of the confusion matrix and stage-wise F1 scores further validate the high recognition rates of the proposed model for the major sleep stages (W, N2, N3, and REM), while also significantly improving the model's discrimination ability for the N1 stage. It is noteworthy that the classification of the N1 stage has long been recognized as a challenging task within the field. In both the Sleep-EDF-20 and Sleep-EDF-78 datasets, several studies have reported F1 scores for N1 typically ranging between 50% and 55% (e.g., 52.2% for SleepEEGNet, 44.4% for DeepSleepNet-lite, and 47.9% for AttnSleep). Compared to our F1 score of 58.4%, this represents a stage that remains a particularly challenging problem with limited room for improvement.

Secondly, the architectural innovations of the model serve as the core driving force behind its performance improvements. Ablation study results demonstrate that the adaptive gating mechanism and channel attention module play a critical role in optimizing the overall performance. The synergistic effect of these components not only enhances the ability to extract multi-time-scale features but also improves the interaction efficiency between feature channels, thereby significantly boosting the model's adaptability to complex EEG signals. Compared to the MultiSEss model, which achieves an Acc of 83.8% and an MF1 score of 73.4% on the Sleep-EDF-20 dataset, ADG-SleepNet clearly exhibits superior performance in overall classification accuracy and the identification of complex stages, despite the former performing well on certain stages. A deeper analysis of the dynamic fusion weights reveals that the model can dynamically adjust the focus range of the multi-dilation branches based on the sleep stages, demonstrating a powerful multi-scale modeling capability and dynamic representation ability. In contrast, while the S4Sleep model shows more balanced F1 scores across multiple stages, particularly in the REM stage with an F1 score of 87.7%, its overall Acc (84.4%) and MF1 (80.7%) are still slightly inferior to ADG-SleepNet's accuracy of 87.1% and macro F1 score of 84.0%. This is particularly evident in the fact that ADG-SleepNet exhibits higher precision and robustness when handling complex sleep stages, such as N1 and N3.

Moreover, ADG-SleepNet achieves high performance while compressing the parameter count to less than 0.49M, which is substantially lower than that of most comparable methods. This lightweight design lays both theoretical and practical foundations for real-time deployment of sleep staging models on wearable and edge devices. The success of this approach indicates that high-quality sleep stage recognition can be realized under resource-constrained conditions through rational architectural innovations and feature fusion mechanisms.

Nevertheless, this study has certain limitations. The model has been primarily validated on single-channel EEG datasets, and its adaptability to multi-channel signals, cross-device data, and complex clinical scenarios warrants further investigation. Additionally, although the model achieves improved performance on difficult stages such as N1, there is still room to enhance its robustness under conditions of extreme class imbalance or

limited sample availability. Future work may further incorporate domain-specific priors, physiological feature augmentation, and multimodal data fusion to continually optimize both the performance and applicability of the model.

5. Conclusions

In this study, we proposed an efficient and lightweight network, ADG-SleepNet, for single-channel EEG-based sleep staging, and systematically validated its effectiveness on two authoritative public datasets: Sleep-EDF-20 and Sleep-EDF-78. The proposed model, with only 0.49 million parameters, achieved overall staging accuracies exceeding 85% and macro-averaged F1 scores above 80%, consistently outperforming current mainstream lightweight models—particularly demonstrating substantial advantages in the challenging N1 stage.

ADG-SleepNet integrates multiple architectural innovations, including multi-dilation parallel branches, adaptive gating, and channel attention mechanisms, to fully exploit the multi-scale and temporal features inherent in sleep EEG signals. Ablation studies and stage-wise analyses further confirm the effectiveness and synergistic contributions of these key modules. Meanwhile, the extremely low parameter complexity of the model provides a solid foundation for its real-time deployment in wearable devices, mobile platforms, and edge computing scenarios.

Overall, ADG-SleepNet achieves an excellent balance between model efficiency and staging accuracy, offering new insights into the development of approaches to lightweight automatic sleep staging. Future work will focus on extending the applicability of the model to more complex scenarios, such as multi-channel data, cross-device environments, and multimodal sleep monitoring, thereby further advancing the practical implementation and innovative application of intelligent sleep health monitoring technologies.

Author Contributions: Conceptualization, H.S. and Z.Z.; methodology, H.S.; software, H.S.; validation, H.S. and Z.Z.; formal analysis, H.S.; investigation, H.S.; resources, Z.Z.; data curation, H.S.; writing—original draft preparation, H.S.; writing—review and editing, H.S. and Z.Z.; visualization, H.S.; supervision, Z.Z.; project administration, Z.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: This study used two publicly available datasets, Sleep-EDF-20 and Sleep-EDF-78, obtained from PhysioNet. All participants in the original studies provided informed consent, and the data-collection protocols were approved by the respective institutional review boards of the original research teams. As we did not recruit new participants or collect new data for this secondary analysis, no additional informed consent forms were generated.

Acknowledgments: The authors would like to thank Hebei GEO University for their valuable technical guidance and support during this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Medic, G.; Wille, M.; Hemels, M.E. Short- and long-term health consequences of sleep disruption. *Nat. Sci. Sleep* **2017**, *9*, 151–161. [\[CrossRef\]](#)
2. Iber, C.; Ancoli-Israel, S.; Chesson, A.L.; Quan, S.F. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, 1st ed.; American Academy of Sleep Medicine: Westchester, IL, USA, 2007.
3. Imtiaz, S.A. A Systematic Review of Sensing Technologies for Wearable Sleep Staging. *Sensors* **2021**, *21*, 1562. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Faust, O.; Hagiwara, Y.; Hong, T.J.; Lih, O.S.; Acharya, U.R. Deep learning for healthcare applications based on physiological signals: A review. *Comput. Methods Programs Biomed.* **2018**, *161*, 1–13. [\[CrossRef\]](#) [\[PubMed\]](#)

5. Phan, H.; Mikkelsen, K. Automatic sleep staging of EEG signals: Recent development, challenges, and future directions. *Physiol. Meas.* **2022**, *43*, 04TR01. [[CrossRef](#)] [[PubMed](#)]
6. Phan, H.; Andreotti, F.; Cooray, N.; Chén, O.Y.; De Vos, M. Automatic sleep stage classification using single-channel EEG: A review. *Sleep Med. Rev.* **2018**, *39*, 78–86. [[CrossRef](#)]
7. Supratak, A.; Dong, H.; Wu, C.; Guo, Y. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 1998–2008. [[CrossRef](#)]
8. Michielli, N.; Acharya, U.R.; Molinari, F. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. *Comput. Biol. Med.* **2019**, *106*, 71–81. [[CrossRef](#)]
9. Boostani, R.; Karimzadeh, F.; Nami, M.A. A comparative review on sleep stage classification methods in patients and healthy individuals. *Comput. Methods Programs Biomed.* **2017**, *140*, 77–91. [[CrossRef](#)]
10. Hassan, A.R.; Bhuiyan, M.I.H. Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. *Comput. Methods Programs Biomed.* **2017**, *140*, 201–210. [[CrossRef](#)]
11. Tsinalis, O.; Matthews, P.M.; Guo, Y. Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Ann. Biomed. Eng.* **2016**, *44*, 1587–1597. [[CrossRef](#)]
12. Tsinalis, O.; Matthews, P.M.; Guo, Y. Automatic sleep stage scoring using raw single-channel EEG with deep convolutional neural networks. *IEEE J. Biomed. Health Inform.* **2016**, *20*, 1253–1262. [[CrossRef](#)]
13. Chambon, S.; Galtier, M.N.; Arnal, P.J.; Wainrib, G.; Gramfort, A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 758–769. [[CrossRef](#)]
14. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
15. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5MB model size. *arXiv* **2016**, arXiv:1602.07360.
16. Han, S.; Pool, J.; Tran, J.; Dally, W.J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *arXiv* **2015**, arXiv:1510.00149.
17. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531.
18. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.
19. Lea, C.; Flynn, M.D.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal convolutional networks for action segmentation and detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 156–165. [[CrossRef](#)]
20. Farha, Y.A.; Gall, J. MS-TCN: Multi-Stage Temporal Convolutional Network for Action Segmentation. *arXiv* **2019**, arXiv:1903.01945.
21. Einizade, A.; Nasiri, S.; Hajipour Sardouie, S.; Clifford, G. ProductGraphSleepNet: Sleep Staging Using Product Spatio-Temporal Graph Learning with Attentive Temporal Aggregation. *Neural Netw.* **2023**, *164*, 667–680. [[CrossRef](#)]
22. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
23. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [[CrossRef](#)]
25. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241. [[CrossRef](#)]
26. Dauphin, Y.N.; Fan, A.; Auli, M.; Grangier, D. Language modeling with gated convolutional networks. In Proceedings of the 34th International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp. 933–941.
27. Srivastava, R.K.; Greff, K.; Schmidhuber, J. Highway networks. In Proceedings of the 28th Conference on Neural Information Processing Systems (NeurIPS), Montreal, QC, Canada, 7–12 December 2015; pp. 1–10.
28. Chen, Y.; Chen, X.; Wang, Y.; Pan, J.; Wang, Y.; Si, Y.; Yang, Y.; Zhang, Y.; Gao, Y. Multi-scale fusion convolutional neural network for automatic sleep stage classification based on single-channel EEG. *Sensors* **2020**, *20*, 6120. [[CrossRef](#)]
29. Eldele, E.; Chen, Z.; Liu, C.; Wu, M.; Kwok, C.K.; Li, X.; Fang, H. An Attention-Based Deep Learning Approach for Sleep Stage Classification With Single-Channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 809–818. [[CrossRef](#)] [[PubMed](#)]
30. Zhao, R.; Zhang, S.; Wang, T.; Zhang, Y.; Guo, Y. NAMRTNet: Nonlocal attention-based multi-resolution temporal convolutional network for sleep stage classification. *Biomed. Signal Process. Control* **2023**, *82*, 104529. [[CrossRef](#)]

31. Olesen, A.N.; Jennum, P.J.; Mignot, E.; Sørensen, H.B.D. Automatic sleep stage classification with deep residual networks in a mixed-cohort setting. *Sleep* **2021**, *44*, zsaa161. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Sors, A.; Bonnet, S.; Mirek, S.; Vercueil, L.; Payen, J.F. A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomed. Signal Process. Control* **2018**, *42*, 107–114. [\[CrossRef\]](#)
33. Kemp, B.; Zwinderman, A.H.; Tuk, B.; Kamphuisen, H.A.C.; Obery, J.J.L. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* **2000**, *47*, 1185–1194. [\[CrossRef\]](#)
34. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [\[CrossRef\]](#)
35. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI), Montreal, QC, Canada, 20–25 August 1995; Volume 2, pp. 1137–1145.
36. Prechelt, L. Early stopping—But when? In *Neural Networks: Tricks of the Trade*; Orr, G.B., Müller, K.R., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 55–69.
37. Mousavi, S.; Afghah, F.; Acharya, U.R. SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach. *PLoS ONE* **2019**, *14*, e0216456. [\[CrossRef\]](#)
38. Fiorillo, L.; Favaro, P.; Faraci, F.D. DeepSleepNet-Lite: A simplified automatic sleep stage scoring model with uncertainty estimates. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 2076–2085. [\[CrossRef\]](#)
39. Seo, H.; Back, S.; Lee, S.; Park, D.; Kim, T.; Lee, K. Intra- and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG. *Biomed. Signal Process. Control* **2020**, *61*, 102037. [\[CrossRef\]](#)
40. Phan, H.; Lorenzen, K.P.; Heremans, E.; Chen, O.Y.; Tran, M.C.; Koch, P.; Mertins, A.; Baumert, M.; Mikkelsen, K.B.; De Vos, M. L-seqsleepnet: Whole-cycle long sequence modelling for automatic sleep staging. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 4950–4961. [\[CrossRef\]](#)
41. Lee, M.; Kwak, H.G.; Kim, H.J.; Won, D.O.; Lee, S.W. SeriesSleepNet: An EEG time series model with partial data augmentation for automatic sleep stage scoring. *Front. Physiol.* **2023**, *14*, 1131050. [\[CrossRef\]](#)
42. Huang, Z.; Yang, Y.; Wang, Z.; Zhang, S. MultiSEss: Automatic sleep staging model based on SE attention mechanism and state-space model. *Biomimetics* **2025**, *10*, 288. [\[CrossRef\]](#)
43. Jirakittayakorn, N.; Wongsawat, Y.; Mitirattanakul, S. ZleepAnlystNet: A novel deep learning model for automatic sleep stage scoring based on single-channel raw EEG data using separating training. *Sci. Rep.* **2024**, *14*, 9859. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Wang, T.; Strodthoff, N. S4Sleep: Elucidating the design space of deep-learning-based sleep stage classification models. *Comput. Biol. Med.* **2025**, *170*, 109735. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Supratak, A.; Guo, Y. TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel EEG. In Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 641–644. [\[CrossRef\]](#)
46. Phan, H.; Mikkelsen, K.; Chen, O.Y.; Koch, P.; Mertins, A.; De Vos, M. SleepTransformer: Automatic sleep staging with interpretability and uncertainty quantification. *IEEE Trans. Biomed. Eng.* **2022**, *69*, 2456–2467. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Li, C.; Qi, Y.; Ding, X.; Zhao, J.; Sang, T.; Lee, M. A deep learning method approach for sleep stage classification with EEG spectrogram. *Int. J. Environ. Res. Public Health* **2022**, *19*, 6143. [\[CrossRef\]](#)
48. Bao, J.; Wang, G.; Wang, T.; Wu, N.; Hu, S.; Lee, W.H.; Lo, S.L.; Yan, X.; Zheng, Y.; Wang, G. A feature fusion model based on temporal convolutional network for automatic sleep staging using single-channel EEG. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 6641–6652. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.