SSC-SleepNet: A Siamese-Based Automatic Sleep Staging Model with Improved N1 Sleep Detection

Songlu Lin, *Student Member, IEEE,* Zhihong Wang, Hans van Gorp, *Student Member, IEEE,* Mengzhu Xu, *Student Member, IEEE,* Merel van Gilst, Sebastiaan Overeem, Jean-Paul Linnartz, *Fellow, IEEE,* Pedro Fonseca, Xi Long, *Senior Member, IEEE*

***Abstract*—Automatic sleep staging from single-channel electroencephalography (EEG) using artificial intelligence (AI) is emerging as an alternative to costly and time-consuming manual scoring using multi-channel polysomnography. However, current AI methods, mainly deep learning models such as convolutional neural network (CNN) and long short-term memory (LSTM), struggle to detect the N1 sleep stage, which is challenging due to its rarity and ambiguous nature compared to other stages. Here we propose SSC-SleepNet, an automatic sleep staging algorithm aimed at improving the learning of N1 sleep. SSC-SleepNet employs a pseudo-Siamese neural network architecture owing to its capability in one- or few-shot learning with contrastive loss. Which we selected due to its strong capability in one- or few-shot learning with a contrastive loss function. SSC-SleepNet consists of two branches of neural networks: a squeeze-and-excitation residual network branch and a CNN-LSTM branch. These two branches are used to generate latent features of the EEG epochs. The adaptive loss function of SSC-SleepNet uses a weighing factor to combine weighted cross-entropy loss and focal loss to specifically address the class imbalance issue inherent in sleep staging. The proposed new loss function dynamically assigns a higher penalty to misclassified N1 sleep stages, which can improve the model's learning capability for this minority class. Four datasets were used for sleep staging experiments. In the Sleep-EDF-SC, Sleep-EDF-X, Sleep Heart Health Study, and Haaglanden Medisch Centrum datasets, SSC-SleepNet achieved macro F1-scores of 84.5%, 89.6%, 89.5%, and 85.4% for all sleep stages, and N1 sleep stage F1-scores of 60.2%, 58.3%, 57.8%, and 55.2%, respectively. Our proposed deep learning model outperformed the most existing models in automatic sleep staging using single-channel EEG signals. In particular, N1 detection performance has been markedly improved compared to the state-of-art models.**

***Index Terms*— Sleep staging, electroencephalography, pseudo-Siamese network, adaptive loss function, N1 sleep.**

I. INTRODUCTION

SLEEP diagnosis is always an important concern for public health and medicinal systems. Polysomnography (PSG) is the gold standard in clinical sleep assessment [1]-[3], which meticulously manually scored following the guidelines of the American Academy of Sleep Medicine (AASM) [4] into one of five distinct sleep stages: wakefulness (W), rapid eye movement (REM) sleep, and three stages of non-rapid eye movement (NREM) sleep (N1, N2, N3) [5]. Despite its comprehensiveness, the use of PSG has its drawbacks [6]; it is often perceived as uncomfortable for patients due to the large number of sensors and wires [7]. In addition, the evaluation process is not only time-consuming but also costly, requiring significant professional involvement, and posing a challenge to widespread clinical use. As a result, there's a shift toward finding simpler alternatives: using more convenient devices and automation to classify sleep stages [8][9][10]. Compared with multi-channel EEG signals, single-channel EEG signals are more easily collected by portable, lightweight devices.

In recent years, deep learning has become the state of the art in EEG-based automatic sleep staging due to its ability to perform automatic feature extraction and achieve superior performance compared to traditional machine learning [11]-[13]. Most current deep learning models for sleep staging rely on convolutional neural networks (CNNs) [12] and recurrent neural networks (RNNs) [14]-[16]. CNNs are effective in extracting features that represent short-term EEG patterns, while RNNs capture long-term temporal dependencies. To enhance this, models often incorporate long short-term memory (LSTM) [15], [17] or multi-head attention mechanisms [18], [19] to further capture temporal dependencies. For example, in 2017, Supratak et al. [20]

Manuscript received 28 April 2024. *(Corresponding author: Xi Long, Zhihong Wang).*

Songlu Lin is with the with the College of Instrument Science and Electrical Engineering, Jilin University, Changchun 130021, China, and also with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: linsl22@mails.jlu.edu.cn).

Zhihong Wang is with the College of Instrument Science and Electrical Engineering, Jilin University, Changchun 130021, China (e-mail: wzh@jlu.edu.cn).

Hans van Gorp, Mengzhu Xu, Merel M. van Gilst, Jean-Paul Linnartz, and Xi Long are with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: x.long@tue.nl).

Sebastiaan Overeem is with the Kempenhaeghe Center for Sleep Medicine, 5591 VE Heeze, and also the Department of Electrical Engineering, Eindhoven University of Technology, 5612AZ Eindhoven, The Netherlands.

Pedro Fonseca is with Philips Sleep and Respiratory Care, 5656 AE Eindhoven, and also the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands.

developed DeepSleepNet based on CNN and bidirectional LSTM networks and achieved an accuracy of 82.0% on the publicly available benchmark dataset Sleep-EDF-SC. Recently, many methods have used attention mechanisms to improve model performance in sleep staging. For example, the SleepEEGNet [21] achieved 84.3% accuracy on Sleep-EDF-SC and 80.0% accuracy on Sleep-EDF-X. The AttnSleep model proposed by Eldele et al. [22] achieved an accuracy of 84.4%, 81.3%, and 84.2% on the Sleep-EDF-SC, the Sleep-EDF-X, and the Sleep Heart Health Study (SHHS) datasets, respectively.

However, existing single-channel EEG models typically follow single-source single-model or serial multi-model structures, which struggle to effectively capture multi-dimensional signal features. Different architectures offer unique strengths: CNNs excel in capturing local features [23]-[25], RNNs model global temporal dependencies [26], [27], and SE-ResNet is particularly adept at enhancing important features while suppressing less relevant ones by explicitly modeling inter-channel dependencies[28]-[30]. A pseudo-Siamese network (PSN) structure [31], [32], processing two different model architectures in parallel, can capture both similarities and differences between these architectures, resulting in a more comprehensive feature extraction process.

N1 is characterized by the loss of alpha activity, emergence of theta waves, and slow rolling eye movements [33], and serves as a transitional phase that may include brief alpha bursts and arousals which plays a key role in sleep architecture and overall sleep quality. However, to the best of our knowledge, all the existing deep learning methods show relatively low performance in detecting N1 sleep [34], [35], with an F1 score of often less than 40-50% depending on the dataset. There can be several reasons for this. Firstly, compared with the other sleep stages, the limited number of N1 epochs might lead to insufficient learning of N1 when training a sleep staging model, as N1 sleep accounts for only approximately 5% of the total sleep time per night [36]. Secondly, the scoring of N1 according to the AASM guidelines relies on EEG patterns that do not occur in all patients (such as the attenuation of alpha rhythm) and on other patterns that appear on EOG but not on EEG (such as slow eye movements) [37], [38]. Finally, the scoring on N1 also relies on the recognition of attenuating or disappearing patterns (like the alpha rhythm); N1 is scored in subsequent epochs until there is evidence for another sleep stage, such as the appearance of K complexes, spindles, etc. These scoring rules, in absentia of patterns, increase ambiguity and lead to substantial disagreement between human scorers [39]-[42]. Rosenberg and Van Hout reported agreement in N1 scoring of 63% against group consensus [43], and in a meta-analysis of inter-rater agreement, Lee et al reported Cohen's kappa of only 0.24 [44]. In turn, this leads to disagreement between sleep staging algorithms and human scorers due to the model not being able to learn the prior knowledge of human scorers well, resulting in judgment errors [45]. In addition, previous research has focused primarily on optimizing overall classification performance (as loss function) for all sleep stages when modeling. The minority N1 stage contributes much less to those overall performance metrics and would likely be poorly optimized or overlooked.

To tackle the challenge of effectively learning from imbalanced datasets with limited N1 sleep stage samples, we employ a Pseudo-Siamese Network (PSN) structure. By deviating from the conventional Siamese model, which relies on complete weight sharing for contrastive learning, PSN introduces architectural flexibility, which means we use the same inputs but different subbranches to share parameters. This allows for partial weight sharing or tailor-made designs for varying inputs, thereby enhancing the model's capacity to robustly represent features by scrutinizing both similarities and disparities across class samples. Leveraging this innovative architecture, we proposed a novel deep-learning framework called SSC-SleepNet for sleep staging with single-channel EEG data. This network utilized two different branches, each with a model that has been well applied to EEG-based sleep staging: a squeeze-and-excitation residual network (SE-ResNet) and a CNN-LSTM network. This enhances feature extraction by learning nuanced features from EEG signals by comparing the two branches. Moreover, we design an adaptive contrastive loss function. This loss function, superior to traditional cross-entropy loss, excels in managing imbalanced datasets while maintaining good classification accuracy. Through this methodology, our framework aims to surpass the limitations of existing models in accurately detecting N1 sleep, providing a more balanced and effective learning approach to sleep stage classification based on single-channel EEG. The main contributions of this study are as follows:

1) We propose a pseudo-Siamese neural network based on a heterogeneous parallel contrastive learning architecture. By extracting intrinsic modality-specific features within different model architectures and establishing joint associations between these modalities, the proposed model maximizes the similarity between features extracted from heterogeneous sub-models while minimizing inter-feature distance, thus improving sleep staging performance.

2) We design an adaptive dynamic loss function, which enhances model training by adjusting weights in response to class imbalances, thereby improving classification accuracy for underrepresented stages like N1.

3) The proposed model shows strong adaptability across various main EEG electrode positions and configurations, consistently achieving good performance on multiple different EEG channels.

The remainder of this paper is as follows: Section II describes SSC-SleepNet details, including PSN structure and the adaptive dynamic loss function. Section III presents experimental methods and settings, Section IV shows the results of the experiments, followed by a discussion in Section V. Finally, Section VI concludes the study.

II. Methods: SSC-SleepNet

A. Overall Framework

The overall framework of the SSC-SleepNet is shown in Fig. 1. For training a sleep staging model, 30-second single-channel

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3572886

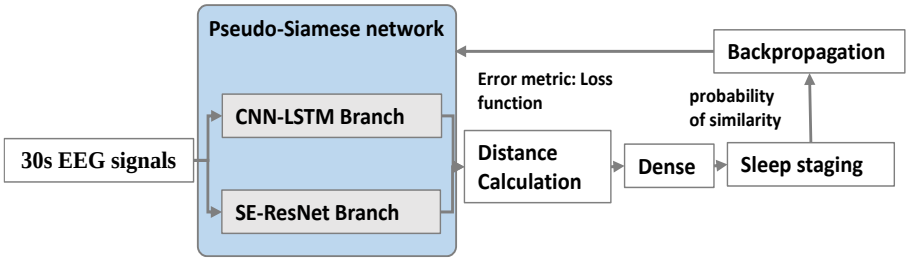

Fig. 1. Overview of the proposed SSC-SleepNet framework.

EEG signals are first preprocessed followed by our proposed PSN. Fig. 2. shows the details of the PSN architecture, including a CNN-LSTM branch and a SE-ResNet branch. These two branches work in parallel and share parameters. The outputs of the two branches are compared in terms of distance before a dense layer for classifying sleep stages. The results can be updated through backpropagation to minimize the loss function.

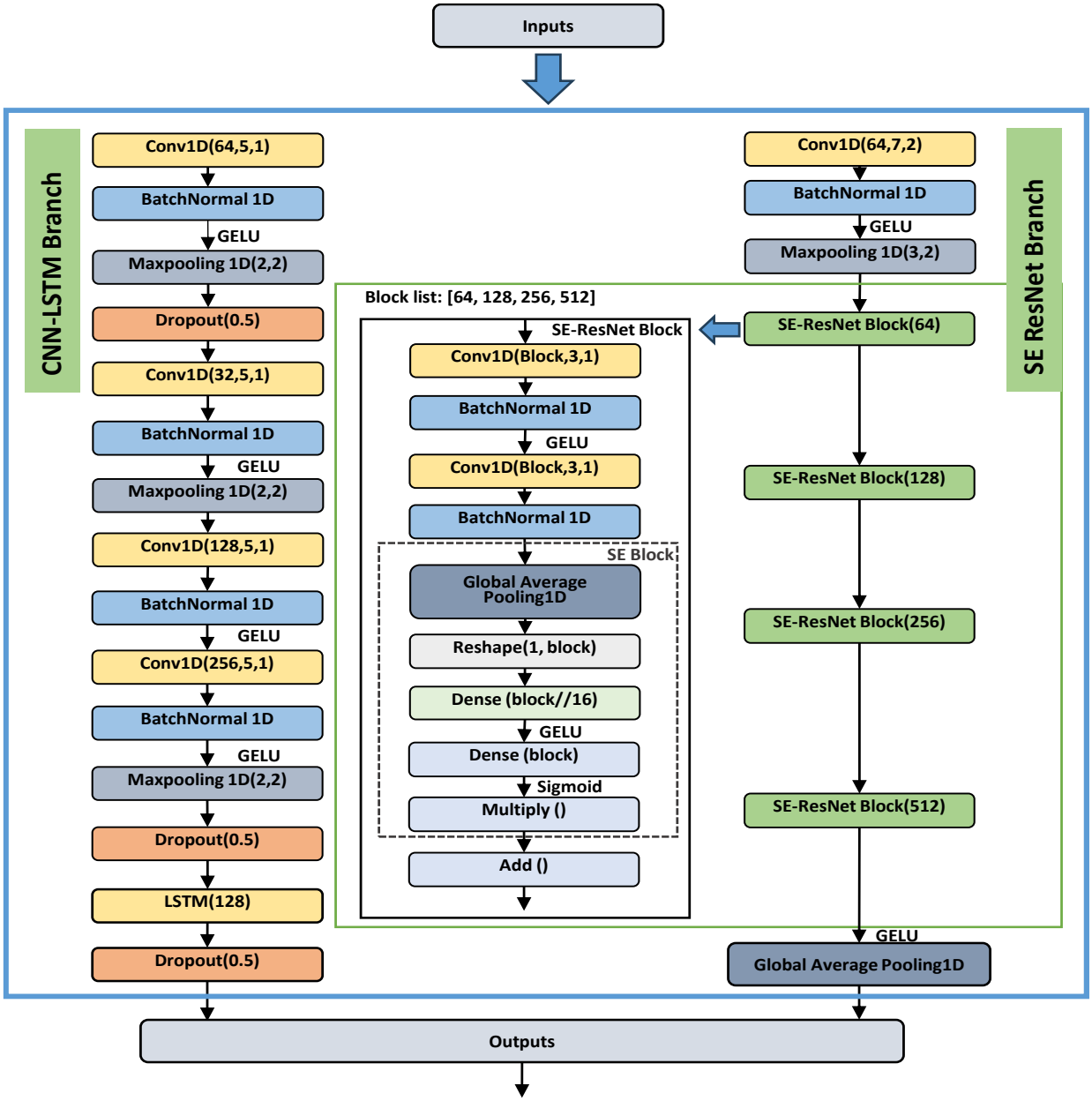

Fig. 2. Architecture of our proposed pseudo-Siamese network.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3572886

B.CNN-LSTM

The CNN-LSTM branch is primarily designed for feature extraction from each 30-second epoch of EEG data. As depicted in Fig. 2. (the left branch), the CNN comprises four convolutional layers. Following each convolutional layer, there is a batch normalization layer, succeeded by an activation function in the form of a Gaussian error linear unit (GELU) [46]. Using the first convolutional layer as an illustration, Conv1D (64, 50, 1) is a one-dimensional convolutional layer indicating 64 channels, a kernel size of 50, and a stride of 1. Likewise, Maxpool1D (2, 2) is a max pooling layer specifying a kernel size of 2 with a stride of 2. Dropout layers with a dropout rate of 0.5 are used to mitigate network overfitting. The feature maps obtained from the CNN layers undergo subsequent processing in an LSTM layer with 128 units.

C. SE-ResNet

The architecture of the SE-ResNet branch is shown in Fig. 2, which consists of an initial convolutional layer, followed by a series of SE-ResNet blocks. The SE-ResNet block is based on a ResNet [47] architecture with an SE part incorporated for performance enhancement [48]. The SE part adaptively recalibrates channel-wise feature responses to emphasize informative features, ultimately improving the representational capacity of the model [49].

The ResNet part of each SE-ResNet block is designed to facilitate the training of deep neural networks by introducing residual connections. This is achieved by adding the input of a given layer to its output, creating a shortcut connection that skips one or more layers. The ResNet consists of a sequence of residual subblocks, each consisting of two or more convolutional layers, followed by batch normalization and activation functions. The final output of the ResNet part is obtained by summing the outputs of the last residual branch with the original network input, forming another residual connection, followed by additional layer normalization.

For SE, global mean pooling is first applied to the input feature map, resulting in a channel-wise descriptor. This descriptor is then processed through two fully connected layers, where the first layer reduces the dimensionality by a factor of *r* (reduction ratio) and the second layer restores the original dimensionality. The output of the second fully connected layer is passed through a sigmoid activation function to produce a set of channel-wise weights. These weights are then used to scale the input feature map, enhancing the representation of important features while suppressing less useful features. This adaptive recalibration process allows the model to focus on the most informative features of the task at hand, thereby improving model performance.

D. Adaptive Contrastive Loss Function

We develop a novel loss function aimed at dynamically boosting the learning of N1 sleep and reducing the negative impact of the N1 sleep stage on the overall performance of automatic sleep staging models. This is achieved by introducing a dynamically generated hyperparameter λ, used to adjust the classification loss during the training process. This method allows for more balanced learning, particularly beneficial for the underrepresented N1 stage. The loss function we designed is:

$$Total\ loss = (1-\lambda) \times J_{WCE} + \lambda \times FL + CL, \quad (1)$$

where J_{WCE} is the weighted cross-entropy loss. FL is the focal loss, and CL is the contrastive loss. λ is a factor dynamically adjusted according to the loss difference between FL and J_{WCE}. Different parts of the loss are described in the following.

1) Weighted cross-entropy loss: Based on existing research [17], the N3 stage has relatively obvious slow wave activity, which can be well classified. Conversely, primarily due to its transitional nature and subtle physiological signs, the distinction between N1 and other stages such as W, N2, and R is the most challenging. As a result, our weighting system assigns the highest weight to N1 and the lowest weight to N3, while it assigns comparable weights to stages such as N2 and R. The function below is the weighted cross-entropy loss (WCE) [50]:

$$WCE = W \times J_{CE}(y,p), \quad (2)$$

where $J_{CE}(y,p) =$

$$-\frac{1}{N}\{\sum_{n=1}^{N}\sum_{c=1}^{C}[y_n^c \log(\hat{y}_n^c) + (1-y_n^c) \times \log(1-\hat{y}_n^c)]\}. \quad (3)$$

In Eq. (3), J_{CE} is the traditional cross-entropy loss function [51], *y* represents the true label, and *p* represents the probability that the model predicts that the sample belongs to each category. *N* is the number of samples. y_n^c and $\hat{y}_n^c$ represent the real and the predicted label, respectively. *W* is a set of class weights.

2) Dynamic focal loss: For unbalanced datasets, Lin T Y et al. [52] proposed a focal loss function to reduce the attention from easily classified samples and focus on difficult samples. The focal loss (FL) can be calculated such that:

$$FL(p_t, \alpha_t, \gamma) = -\alpha_t(1-p_t)^{\gamma}\log(p_t), \quad (4)$$

where p_t is the probability of model prediction, α_t is the set of class weights, and γ is the adjustment factor. As γ increases, the model pays less attention to the "easy-to-classify" samples but more to those "difficult-to-classify" samples. *t* represents the true label of each sleep stage, α_t and γ can be automatically adjusted based on the model performance.

3) Contrastive loss: The distance function *D* between the first branch inputs x_1 and the second branch inputs x_2 is defined as the Euclidean distance [53] :

$$D(x_1, x_2) = ||\mathrm{y}(x_1)_i - \mathrm{y}(x_2)_i||_2, \quad (5)$$

where $\mathrm{y}(x_1)_i$ represents the i^{th} element output of the first branch, and $\mathrm{y}(x_2)_i$ represents the i^{th} element output of the second branch. The contrastive loss [54] of the PSN is:

$$CL(y,d,m) = \frac{1}{2N} y_i D_i^2 + (1-y_i)\max(0, m-D_i)^2, \quad (6)$$

where *N* represents the number of samples, y_i represents a binary label used to indicate whether a pair of samples (from the two branches) belongs to the same class. If two samples belong to the same category, then $y_i = 1$; otherwise $y_i = 0$. D_i represents $D(x_1, x_2)_i$, and m is a pre-defined boundary value used to determine when samples are considered similar.

As stated, we use the CNN-LSTM model and the SE-ResNet model as the two branches of the PSN. During training, three

hyperparameters are used to adaptively adjust the loss function. In the focal loss function (In Eq.(4).), we adjust α and γ according to the classification accuracy after each training. The accuracy is initially set to 0.7. When the accuracy is higher than the last training result, a step size value is added to α and γ; while if lower, they are reduced by the same step size values. The range of α was set to 0.1–1.0 with a step size of 0.05 and the range of γ was 0–5.0 with a step size of 1.The initial value of α set 0.25 and the initial value of γ set 2.0. In addition, in Eq. (1), λ is dynamically adjusted according to the loss difference between FL and J_{WCE}. If the J_{WCE} value is larger than the FL value, then the value of λ will be increased. The range of λ ranged from 0.01 to 0.99 with a step size of 0.05 and an initial value of 0.75.

TABLE I
DETAILS OF THREE DATASETS USED IN OUR EXPERIMENTS

Datasets	Channel	Subjects	Recordings	Sampling Rate	W	N1	N2	N3	R	Total
Sleep-EDF-SC	Fpz-Cz/Pz-Oz	20	39	100 Hz	8285	2804	17799	5703	7717	42308
Sleep-EDF-X	Fpz-Cz/Pz-Oz	78	153	100 Hz	65951	21522	69132	13039	25835	195479
SHHS	C4-A1	329 5463	329 5463	125 Hz 125 Hz	46319 445627	10304 61898	142125 665508	60153 222570	65953 241922	324854 1637525
HMC	F4-M1/C4-M1/O2-M1/C3-M2	151	151	256 Hz	23686	15548	50083	26671	21255	137243

Note that, W, N1, N2, N3 and R indicate the number of 30-second epochs for each sleep stage, on each dataset.

III. Methods: Experiments and Evaluation

A. Datasets

In this study, as shown in Table I, we adopted four PSG datasets, namely Sleep-EDF-SC [55], [56], Sleep-EDF-X [57], SHHS [58], [59], and Haaglanden Medisch Centrum sleep staging dataset (HMC) [60] to evaluate the effectiveness of the proposed SSC-SleepNet for sleep staging based on single-channel EEG signals. The details and the use of each dataset are provided in the following.

1) Sleep-EDF-SC dataset: Originating from the Sleep-EDF database in PhysioNet, the Sleep-EDF-SC dataset consists of 39 overnight PSG recordings from 20 healthy subjects aged 25-34 years, with an equal gender distribution. This dataset is divided into two subsets: Sleep Cassette (SC) and Sleep Telemetry (ST). While the Sleep Cassette subset examines the effects of age on sleep, the Sleep Telemetry subset examines the effects of temazepam. The Sleep-EDF-SC participants wore a modified Walkman-like cassette-tape recorder [61]. The recorded data include EEG (Fpz-Cz channel and Pz-Oz channel) , EOG, and chin EMG signals sampled at 100 Hz and event markers. Some records also contain respiration and body temperature. For our investigation, we specifically used 39 recordings of the single EEG channel (Fpz-Cz channel or Pz-Oz channel) from the Sleep Cassette subset.

2) Sleep-EDF-X dataset: Sleep-EDF-X, an extension of Sleep-EDF-SC, presents 153 overnight PSG recordings. These recordings include data from 20 healthy participants and 58 patients diagnosed with mild sleep difficulties. This spectrum ensures coverage of both conventional and slightly atypical sleep patterns. Participants range in age from 25 to 101 years old and include 41 males and 37 females. For each recording, a spectrum of physiological signals was recorded, including EEG, EOG, and EMG. Sleep stages were scored per epoch according to the Rechtschaffen and Kales (R&K) criteria [62] as one of eight established categories: "W", "S1", "S2", "S3", "S4", "R", "Movement", and "Unscored".

3) SHHS dataset: The SHHS is a multicenter cohort study focusing on the cardiovascular consequences of sleep-disordered breathing. SHHS participants underwent an overnight in-home PSG using the Compu medics Portable PS-2 System (Abbotts Ville, Victoria, Australia) administered by trained technicians. The recordings include C3-A2 and C4-A1 EEGs sampled at 125Hz, and some other physiological signals such as EOGs for both the right and left eyes, chin EMG , and ECG sampled at 125Hz. Subjects enrolled in the study have a variety of conditions, ranging from pulmonary and cardiovascular diseases to coronary complications. Of the SHHS dataset, we considered two subsets in this work. First, we considered all SHHS data (Visit-1) to evaluate SSC-SleepNet model performance, similar to previous studies [23]. Note that the recordings which did not have all five sleep stages (about 2% of the entire database) were excluded, resulting in 5,463 recordings. Second, to minimize the impact of these diseases and compare the prediction differences between OSA patients and non-OSA patients as done in previous studies [20], [22], subjects with sleep disordered breathing (with an Apnea-Hypopnea Index less than 5) were excluded. This led to a subset of 329 subjects from the SHHS dataset. For these two datasets, we used the C4-A1 EEG channel, recorded at a sampling rate of 125 Hz.

4) HMC dataset: The Haaglanden Medisch Centrum in The Hague, the Netherlands, collected a diverse dataset of 151 whole-night PSG sleep recordings in 2018. This collection used the SOMNO screen Plus and 10-20 (SOMNO medics Germany) recording device and includes data from 85 male and 66 female participants with an age of 53.9 ± 15.4 years (mean ± standard deviation). The data consists of four EEG (F4-M1, C4-M1, O2-M1, and C3-M2), two EOG (E1-M2 and E2-M2), one bipolar chin EMG, and one ECG (single modified lead II) derivations. This montage meets the minimal recommended technical specifications for visual scoring of sleep stages according to the 2.4. version of the AASM guidelines. For our analysis, we used the single EEG channel from the recordings, which are sampled at a frequency of 256 Hz.

For the data preprocessing, we were guided by the methodologies enumerated in previous works [21], [22], [63], [64]. The preprocessing protocols for all datasets involved segmenting the recordings of EEG signals into 30-second epochs to ensure synchronization with the accompanying labels,

suggested by the AASM. Epochs labeled as "movement" and "unscored (signed ?)" were excluded. To use consistent scoring labels, we merged the “S3” and “S4” from the R&K scorings in Sleep-EDF and SHHS datasets into the "N3" AASM sleep label. To sharpen our analysis of distinct sleep cycles, our study only included the 30 minutes of wakefulness preceding the first sleep epoch of the recording [20]. As different datasets have different acquisition channels and acquisition devices, we cannot solve the problem of inconsistent acquisition channels for different datasets. Therefore, we selected different EEG channels of the same dataset for horizontal comparison experiments in Sleep-EDF-SC, Sleep-EDF-X and HMC.

B. Experimental Design

To evaluate the performance of various models, We used k-fold cross-validation [65] to evaluate our model. Given a dataset with N_d subjects, during each fold, the recordings of $N_d - (N_d/k)$ subjects serve as the training set, while recordings from the remaining (N_d/k) subjects were used for validation. This subject-level split guarantees that for a given cross-validation iteration, all data of a given subject is only part of the training or of the validation splits, but never of both. Related to the relatively small size of the Sleep-EDF-SC, Sleep-EDF-X, SHHS-329, and HMC datasets, we consistently set the value of *k* at 20, similar to previous studies [21], [23], [64]. The cross-validation procedure allows us to assess the performance over the complete dataset. After repeating the process *k* times, each time selecting different subjects for the test set, the predicted results from all test sets across all folds were combined to provide a comprehensive assessment across all subjects in the dataset.

We also randomly split the subject into 70% for training and 30% for testing in SHHS-5463, referring to the pervious works [11], [66]. From the training set, 100 subjects were used for validation, same as the studied done previously [11], [67].

C. Model Comparison and Evaluation Metrics

1) Ablation experiments: To investigate the impact of different components within the proposed SSC-SleepNet framework on sleep staging, ablation experiments are performed on the Sleep-EDF-SC.

First, we obtained several models by removing different network components of the SSC-SleepNet and applying the same adaptive contrastive loss functions. These models and their configurations are as follows:

a. CNN-LSTM: a model including only CNN and LSTM and using adaptive dynamic loss function.
b. SE-ResNet: a model combining SE and ResNet and using adaptive dynamic loss function.
c. Siam-CNN-LSTM: a Siamese network model including two branches of CNN-LSTM blocks and using adaptive dynamic loss function.
d. Siam-SE-ResNet: a Siamese model including two branches of SE-ResNet blocks and using adaptive dynamic loss function.
e. SSC-SleepNet: our proposed model, including one branch of CNN-LSTM and one of SE-ResNet and using adaptive dynamic loss function.

Besides, we also compared six loss functions of SSC-SleepNet in Sleep-EDF-SC dataset, including the weighted cross-entropy loss (WCE), the focal loss (FL), the contrastive loss (CL), the combination of the weighted cross-entropy loss and the focal loss (WCE+FL), the combination of the weighted cross-entropy loss and the contrastive loss (WCE+CL) and our proposed adaptive dynamic loss function that combines WCE, FL and (WCE+FL+CL). Here the factor λ was experimentally set to 0.75. This ablation experiment was conducted to understand the impact of each loss on the final sleep staging performance.

2) Evaluation metrics: To assess the classification performance per class, we employed precision (PR), recall (RE), F1-score (F1), and G-mean (GM) as metrics. Moreover, accuracy (ACC), macro F1-score (MF1), Cohen's Kappa (κ) [68], and macro-averaged G-mean (MGm) are utilized to evaluate the overall classification performance. We evaluate these by aggregating all epochs from the test recordings. We denote the true positive, false positive, true negative, and false negative of ith class by TP_i, FP_i, TN_i, and FN_i respectively. The formulas for PR, RE, F1, ACC, MF1, and MGm are then expressed as follows:

$$PR_i = \frac{TP_i}{TP_i+FP_i} \tag{7}$$

$$RE_i = \frac{TP_i}{TP_i+FN_i} \tag{8}$$

$$F1_i = \frac{2*PR_i*RE_i}{PR_i+RE_i} \tag{9}$$

$$ACC = \frac{\sum_{i=1}^{C} TP_i}{S} \tag{10}$$

$$MF1 = \frac{1}{C}\sum_{i=1}^{C} F1_i \tag{11}$$

$$MGm = \frac{1}{C}\sum_{i=1}^{C} \sqrt{\frac{TN_i}{TN_i+FP_i} * \frac{TP_i}{TP_i+FN_i}}, \tag{12}$$

where S represents the total number of samples and C represents the total number of classes.

3) Model performance under N1 optimization: Given that our approach prioritizes improving the performance of the N1 sleep stage while minimizing the impact on other metrics, we conducted extreme bias experiments targeting N1 within the model. This was driven by our desire to uncover avenues for even better predictions of the N1 sleep stage. Thus, we shifted the focus of the loss function from the original global accuracy to exclusively targeting N1 (thereby emphasizing N1 importance).

4) Visual Representation of Features: The automatic classification results obtained with the SSC-SleepNet were compared with the reference from manual scoring of the PSG, available for each dataset. In order to better explore the representation learning process of SSC-SleepNet, we used the data analysis method of uniform manifold approximation and projection (UMAP) [69] to visualize the dimensionality-reduced features of the raw EEG signal, CNN and LSTM branches, combining CNN and LSTM and SE-ResNet, and the third-to-last fully connected layer.

5) Performance Comparison: We compared our method with existing state-of-the-art methods. First, we evaluated the model's performance against other state-of-the-art models by

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3572886

comparing metrics such as total epochs, F1 score, ACC, MF1, MGm, Kappa, and average training time for each sleep stage across several datasets, including Sleep-EDF-SC, Sleep-EDF-X, SHHS-5463, SHHS-329, and HMC. Second, we compared model parameters and floating-point operations (FLOPs).

D. Statistical Analysis

We also calculated sleep statistical indicators (mean ± standard deviation) for each recording, including total wake time (TWT), total sleep time (TST), sleep efficiency (SE), wake after sleep onset (WASO), and sleep stages (N1, N2, N3, R), and used Bland Altman plots [70] to assess the bias (mean error) and the 95% limits of agreement (bias plus/minus 1.96 times the standard deviation, std) of the differences between sleep statistics (SE, WASO and TST) obtained from manual annotation and from automatic scoring with our method.

E. Parameter Optimization

To enable parameter optimization, training, and evaluation of models, we chose Python 3.10 as the programming language and TensorFlow 2.13.0 as the deep learning framework. Two NVIDIA RTX A5000 GPUs were used to perform computations. We used the "*ReduceLROnPlateau*" method [71] to adjust the learning rate during training. Additionally, we dynamically adjusted parameters α, γ, and λ in the proposed loss function (as described in *Section II-E*) to optimize the model's performance. The "*Adam*" optimizer [72] was applied to minimize our class-aware loss and learn the model weights.

IV. Results

Fig. 3. illustrates the comparison of the classification results between the five model variations in the ablation experiments. The overall classification performance of PSN models is better than using the SE-Resnet branch or the CNN branch alone, which further confirms the effectiveness of fusing global hierarchical features and local time features.

We also performed ablation experiments on the loss functions of SSC-SleepNet. Fig. 4. shows the performance of SSC-SleepNet with different loss functions: WCE, FL, CL, WCE+FL, WCE+CL and WCE+FL+CL. The adaptive dynamic loss function (WCE+FL+CL) achieved the best results for all metrics, including ACC, MF1, mGm and Kappa. Notably, the WCE+FL combination outperformed WCE+CL, highlighting the effectiveness of the academic parameter λ in

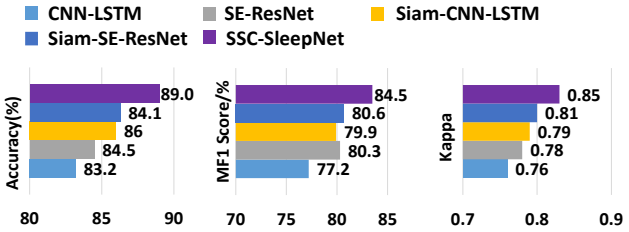

Fig. 3. Ablation experiment on Sleep-EDF-SC.

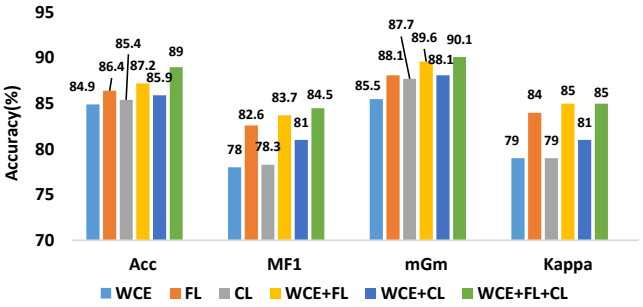

Fig. 4. The ablation for different loss functions from Sleep-EDF-SC.

TABLE II
CONFUSION MATRIX AND PER-CLASS METRICS OF SSC-SLEEPNET SLEEP STAGING MODEL ON THE SLEEP-EDF-SC IN FPZ-CZ EEG CHANNEL

	Predicted					Per-class Metrics			
	W	N1	N2	N3	R	RE (%)	PR (%)	F1 (%)	GM (%)
W	**7515**	313	158	15	184	91.8	95.5	93.6	95.3
N1	222	**1674**	328	4	576	59.7	60.8	60.2	76.2
N2	46	249	**16594**	382	528	93.2	91.6	92.4	93.5
N3	19	2	510	**5165**	7	90.6	92.8	91.6	94.6
R	68	517	532	3	**6597**	85.5	83.6	84.5	90.7

TABLE III
CONFUSION MATRIX AND PER-CLASS METRICS OF SSC-SLEEPNET SLEEP STAGING MODEL ON THE SLEEP-EDF-X

	Predicted					Per-class Metrics			
	W	N1	N2	N3	R	RE (%)	PR (%)	F1 (%)	GM (%)
W	**62911**	1280	722	54	984	95.4	90.6	93.0	98.6
N1	2118	**11690**	4308	525	2881	54.3	63.0	58.3	73.1
N2	3686	2515	**58201**	1634	3096	84.2	88.6	86.4	90.9
N3	17	65	948	**11968**	41	91.8	83.2	87.3	96.1
R	668	3004	1484	197	**20482**	79.3	74.5	76.8	89.4

TABLE IV
CONFUSION MATRIX AND PER-CLASS METRICS OF SSC-SLEEPNET SLEEP STAGING MODEL ON THE SHHS-5643 IN C4-A1 EEG CHANNEL

	Predicted					Per-class Metrics			
	W	N1	N2	N3	R	RE (%)	PR (%)	F1 (%)	GM (%)
W	**413306**	7325	18406	1688	4902	92.8	94.2	93.5	95.3
N1	9762	**32939**	9665	1261	8271	53.2	63.1	57.8	72.5
N2	9102	9043	**611504**	18931	16928	91.9	90.2	91.0	92.5
N3	1406	17	19002	**201365**	780	90.5	89.7	90.1	94.3
R	4996	2857	19579	1143	**213347**	88.2	87.4	87.8	92.9

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3572886

TABLE V
CONFUSION MATRIX AND PER-CLASS METRICS OF SSC-SLEEPNET SLEEP STAGING MODEL ON THE SHHS- 329 IN C4-A1 EEG CHANNEL

	Predicted					Per-class Metrics			
	W	N1	N2	N3	R	RE (%)	PR (%)	F1 (%)	GM (%)
W	**41127**	385	2081	202	2524	88.8	84.9	86.8	93.0
N1	1509	**5528**	1294	216	2257	51.2	73.9	60.5	71.3
N2	2795	741	**124184**	4148	10257	87.4	89.1	88.3	89.5
N3	309	170	5312	**53860**	502	89.5	92.0	90.8	93.8
R	2721	661	6436	115	**56020**	84.9	78.3	81.5	89.4

TABLE VI
CONFUSION MATRIX AND PER-CLASS METRICS OF SSC-SLEEPNET SLEEP STAGING MODEL ON THE HMC IN F4-M1 EEG CHANNEL

	Predicted					Per-class Metrics			
	W	N1	N2	N3	R	RE (%)	PR (%)	F1 (%)	GM (%)
W	**19581**	2328	825	28	924	82.7	84.1	83.4	89.4
N1	2166	**8493**	2080	719	2080	54.7	55.6	55.2	71.9
N2	1060	2812	**39704**	2490	4017	79.3	86.1	82.5	85.7
N3	54	139	2119	**23885**	474	89.6	85.8	87.6	92.9
R	425	1492	1399	726	**17213**	81.0	69.7	75.0	87.0

TABLE VII
CONFUSION MATRIX AND PER-CLASS METRICS OF THE SSC-SLEEPNET UNDER N1 OPTIMIZATION USING THE SLEEP-EDF-SC IN FPZ-CZ EEG CHANNEL

	Predicted					Per-class Metrics			
	W	N1	N2	N3	R	RE (%)	PR (%)	F1 (%)	GM (%)
W	**7443**	455	109	68	210	89.8	88.1	89.0	93.4
N1	193	**2179**	173	7	252	77.7	59.4	**67.3**	86.5
N2	337	483	**15308**	642	1029	86.0	90.7	88.3	89.7
N3	68	10	671	**4945**	9	86.7	85.7	86.2	92.1
R	404	541	610	109	**6053**	78.4	80.1	79.3	86.6

balancing the contributions of focal and contrastive losses. The superior performance of WCE+FL+CL is due to the complementary strengths of these loss functions, which balance data representation, emphasis difficult cases, and enhance feature discrimination, resulting in optimal performance across all metrics.

Tables II-VI show the evaluation results of the SSC-SleepNet model on the Sleep-EDF-SC, Sleep-EDF-X, SHHS-5463, SHHS-329 and HMC in Fpz-Cz, Fpz-Cz, C4-A1, and F4-M1 EEG channel. The results include the confusion matrix, as well as the per-class evaluation metrics, including PR, RE, and F1. In our experiments, the rows of the confusion matrix represent the true labels, and the columns represent the predicted labels. The results of Pz-Oz channel of Sleep-EDF-SC and Sleep-EDF-X, C4-M1, O2-M1, C3-M2 EEG channel of HMC is shown in Table IX. The SSC- SleepNet shows high performance across all metrics in the classification of "W", "N1", "N2", "N3", and "R" sleep stages in all four datasets.

Table VII shows the results obtained in the Sleep-EDF-SC dataset after optimizing the model for N1 detection. The F1 score increases from 61.0% to 67.3%, albeit with a concurrent but slight decrease in performance for the other sleep stages.

We visualize the feature distribution of each component of SSC-SleepNet on the Sleep-EDF-SC Fig. 5. The features of the five sleep stages in the raw EEG signals appear disorganized and lack clear distinguishability. we opt for the same set of parameters to generate UMAP, mainly for the purpose of fair and consistent comparisons across the datasets. The features output by the CNN and LSTM branch and the Se-Resnet branch produce a relatively clear preliminary discrimination among the four classes. Ultimately, compared with SiamSleepNet (see Fig. 5c), after adding the dynamic loss function, the final fully connected layer can more clearly separate sleep stages.

A comparison of the resulting overnight sleep statistics for both our SSC-SleepNet and manual scoring is shown in Table VIII. In this table, we show the mean and standard deviation as

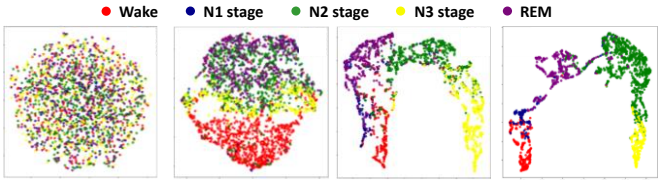

Fig. 5. UMAP visualization of latent features in the representation learning process of SSC-SleepNet on the Sleep-EDF-SC. (a) Raw EEG epochs. (b) Outputs of the CNN and LSTM branch. (c) Outputs of combining CNN and LSTM and SE-ResNet. (d) Outputs of the third-to-last fully connected layer.

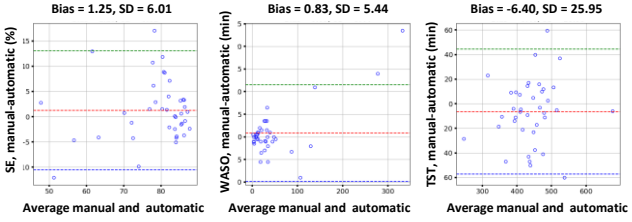

Fig. 6. Bland-Altman graphs for each sleep index in Sleep-EDF-SC. (a) SE, (b) WASO, and (c) TST. The X axis and Y axis represent the mean of the measurements of the experts and those of the proposed method and the difference between the results of the manual and the automatic method, respectively. The horizontal lines in the Bland-Altman plots represent the mean (bias = mean error, 95% limits of agreement = 1.96* standard deviation of the error).

calculated over the recordings. Due to our unified data processing, the TWT, TST, WASO, and SE in the four datasets

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3572886

TABLE VIII
SLEEP MEASURES SUMMARY ASSESSED WITH MANUAL ANNOTATION AND SSC-SLEEPNET ANNOTATION (MEANS ± STANDARD DEVIATIONS)

Dataset	Sleep-EDF-SC		Sleep-EDF-X		SHHS (N=329)		HMC	
Annotation	**Manual**	**Automatic**	**Manual**	**Automatic**	**Manual**	**Automatic**	**Manual**	**Automatic**
TWT(min) *	130.7±118.9	145.5±82.2	228.2±187.7	224.2±128.8	89.9±24.8	95.8±12.0	78.4±64.0	81.9±29.7
TST (min) *	436.2±71.6	426.0±65.2	423.3±62.1	428.8±64.4	423.3±31.1	428.3±28.4	376.0±87.5	384.8±72.2
SE (%)*	79.0±10.6	74.8±9.3	69.0±15.4	64.6±6.3	82.5±4.7	81.9±2.1	82.8±13.9	81.0±6.56
WASO (min)	46.2±70.2	48.7±41.1	39.1±46.0	42.0±48.3	35.4±19.5	35.8±12.0	60.0±53.6	61.4±44.5
N1 (min)	36.0±21.0	27.3±25.3	70.3±45.8	70.4±41.0	15.7±9.7	18.8±25.3	51.5±34.4	55.5±14.4
N2 (min)	228.2±65.9	224.2±44.9	225.9±56.3	235.6±22.6	216.0±42.0	218.5±37.3	165.8±51.2	171.3±50.5
N3 (min)	73.1±36.7	69.2±29.7	42.6±40.4	38.9±13.6	91.4±39.1	82.3±38.4	88.3±42.2	85.9±40.5
R (min)	98.9±27.9	105.3±33.2	84.4±31.4	83.9±18.4	100.2±17.6	108.8±16.8	70.4±31.6	72.1± 28.7

Abbreviations: TWT — total wake time; TST — total sleep time; SE — sleep efficiency; WASO — wake after sleep onset; N1 — sleep stage 1; N2 — sleep stage 2; N3 — sleep stage 3; and R — rapid eye movement sleep stage.

* Due to we did the unified data processing, the TWT, TST, SE and WASO in four datasets are not representing the real-life scenario, but simply for comparison.

do not reflect the real-life scenario but are provided here for the sake of comparison. It can be observed that the estimated overnight sleep statistics using our proposed methodology are very close to those calculated from the human scorers, especially for the time spent in the N1 stage.

The results of the Bland-Altman graph, in Fig.6., showed that the sleep indices obtained using the manual scoring and the proposed method were similar in terms of the SE, WASO and TST.

In order to assess the performance of the proposed SSC-SleepNet in sleep stage classification relative to other state-of-the-art models, we conducted comparative experiments, and the detailed results are presented in Table IX. SSC-SleepNet can distinguish W, N1, and N2 sleep stages well, but N3 and REM might be lower than other models' performance, which may be because we concentrate more on N1 and less attention is paid to N3 and REM. SSC-SleepNet also achieves the highest accuracy, MF1 score, and kappa score. To compare the computational complexity of different methods, Table X shows the number of parameters as well as the FLOPs. The calculation complexity of AttnSleep, with 4.6M parameters, is 33.1M FLOPs. The application of the adopted dynamic loss increased the computational complexity, where multiple autocorrelation calculations on the input data were required.

TABLE X
COMPREHENSIVE COMPARISON OF EFFICIENCY AND PERFORMANCE OF THE STATE-OF-ART MODELS

Model	Parameters (M)	FLOPs (M)
DeepSleepNet** [81] [20]	~24.7	/
DeepSleepNet+** [85] [86]	~24.8	**~55.0**
DeepSleepNet-Lite* [81]	~0.6	/
XSleepNet2** [81] [63]	~5.8	/
SleepEEGNet** [81] [21]	~2.6	/
SeqSleepNet** [67] [81]	~0.2	
U-Sleep* [83]	~3.1	/
SleepTransformer* [84]	~3.7	/
AttnSleep**[85] [22]	~0.5	~64.0
SSC-SleepNet (ours)	~4.6	~33.1

Note that, in all models, * represents the results reported in previous works, are not compatible for a direct comparison here due to the discrepancies in data split, the number of channel used, modelling tasks, etc. ** represents the results reported on second previous works, which may cause different results due to the different environments or parameters, etc., the first [ref.] represents the second report work and the second [ref.] represents the original work reports.

V. DISCUSSION

In our approach, we used a novel pseudo-Siamese neural network model to learn hierarchical and temporal sleep stage features. This enables the model to extract and learn complex representations and patterns from a single EEG signal across different temporal and spatial scales. We introduced an adaptive contrastive loss function that allows the model to dynamically adapt to unbalanced sample size problems to improve the overall classification performance. The significance of this study lies in its accurate prediction of the N1 sleep stage, an indicator of overall sleepiness and sleep quality [74], [75].

A. Pseudo-Siamese Neural Network

The Pseudo-Siamese Neural Network, inspired by contrastive learning, uses a dual-branch network structure, aiming to maximize the predictive accuracy of the model by calculating the similarity between the outputs of these two branches. Here we choose two separates' models, one for each branch. For the first branch, we adopt a CNN-LSTM model. These models are widely used in the field of sleep staging [76]-[78], effectively learning both local features and temporal sequences from electroencephalogram signals. For the second branch, we adopt the SE-ResNet model which has been widely applied in image recognition in recent years. This model is capable of learning critical optimal features at deeper network layers, thus overcoming the problems of overfitting and degradation in deep network models. It also excels in learning global deep-layer feature patterns. By integrating these two network models into a unified framework, we developed a network capable of learning both local temporal features and global deep-layer features.

B. Improvement of the Loss Function

Designing an effective loss function is critical for building and improving deep learning models. A polynomial loss function can be used to improve model performance and generalization by adjusting its expansion coefficients [79]. Therefore, we incorporate these coefficients to improve the traditional cross-entropy loss function. In the presence of uneven distribution of different sleep stages, traditional cross-entropy may bias the model towards predicting more common categories, potentially leading to biased classification results. Therefore, we have integrated the focal loss function into the SSC-SleepNet model to dynamically adjust the model's predictions. The dynamic parameters α and γ in focal loss can focus on the weighting of small sample data sets based on the model's predictions. As a result, this approach achieves higher F1 scores for the minority class N1 when compared with approaches reported in literature, on the same data. Additionally, we introduced a dynamic variable λ to dynamically weigh the results of focal loss in the adaptive

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3572886

contrastive loss function, ensuring that the overall performance is not compromised. This result further validates the effectiveness of the designed adaptive contrastive loss function in improving model performance and mitigating data imbalance.

C. EEG Channels

In clinical sleep research, PSG is widely used as a comprehensive method to assess sleep, integrating a wide range of physiological signals. In our study of analyzing the HMC dataset, we found that the F4-M1 channel demonstrated performance in N1 sleep stage classification. This may be attributed to the enhanced ability of the prefrontal region to capture specific EEG features, such as the α-to-θ transition and prominent slow waves. Notably, previous studies have shown that the AASM-recommended EEG montage, which uses F4-M1 as the primary sensor, achieves higher inter- and intra-rater agreement for N1 scoring compared to the "acceptable" montage configurations, such as C4-M1 [66]. This supports our findings, suggesting that frontal referential montages offer clearer distinguishability of N1 sleep stages compared to frontal bipolar and central referential electrode configurations.

D. Staging Sleep Staging in Patients with Sleep Disorders

In addition to health subjects, we evaluated SSC-SleepNet on the SHHS-5463 and HMC datasets to assess automatic sleep staging in patients with sleep disorders, primarily with OSA. In SHHS, we compared the sleep staging performance of SSC-SleepNet for non-OSA patients (AHI < 5) with that for a general

TABLE IX
COMPARISON WITH STATE-OF-THE-ART MODELS ON SLEEP-EDF-SC, SLEEP-EDF-X, SHHS-5643, SHHS-329 , AND HMC DATASETS.

Dataset	Method	Channel(s)	# 30-s epoch	Per-class F1					Overall Metrics				Average training time / fold
				W (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)	Acc (%)	MF1 (%)	MGm (%)	κ	
Sleep-EDF-SC	DeepSleepNet* [20]	Fpz-Cz	41950	84.7	46.6	85.9	84.8	82.4	82.0	76.9	-	0.76	2.5 hrs**
	DeepSleepNet** [80] [20]	Fpz-Cz	42308	86.7	45.5	85.1	83.3	82.6	82.0	76.9	-	0.76	-
	DeepSleepNet** [64] [20]	Fpz-Cz	42308	86.0	45.0	85.1	84.0	82.6	82.0	76.9	-	0.76	-
	DeepSleepNet-Lite* [81]	Fpz-Cz	42308	87.1	44.4	87.9	88.2	82.4	84.0	78.0	-	0.78	-
	ResAtten* [82]	Fpz-Cz	42308	90.2	48.3	87.8	85.6	83.0	84.3	79.0	-	0.78	-
	AttnSleep* [22]	Fpz-Cz	42308	89.7	42.6	88.8	90.2	79.0	84.4	78.1	85.5	0.79	**21 min**
	U-Sleep* [83]	Majority vote	-	93.0	57.0	86.0	71.0	**88.0**	-	79.0	-	-	-
	SeqSleepNet**[22] [23]	Fpz-Cz	42308	87.7	43.8	88.2	86.5	84.0	84.6	78.0	85.3	0.79	2.5hrs**
	SeqSleepNet**[80] [23]	Fpz-Cz	42308	91.2	44.7	88.0	86.2	83.0	85.6	78.6	-	0.80	-
	L-SeqSleepNet* [80]	Fpz-Cz	42308	91.6	45.3	88.5	86.2	85.2	86.3	79.3	-	0.81	-
	XSleepNet1* [63]	Fpz-Cz EEG, EOG	42308	91.3	49.5	88.0	86.9	84.2	86.0	80.0	-	0.81	-
	XSleepNet2* [63]	Fpz-Cz EEG, EOG	42308	92.2	51.8	88.0	86.8	83.9	86.3	80.6	-	0.81	-
	SleepEEGNet* [21]	Fpz-Cz	42308	89.2	52.2	86.8	85.1	85.0	84.3	79.8	-	0.79	1.5 hrs**
	LGSleepNet* [64]	Fpz-Cz	42308	91.8	49.4	89.6	89.8	82.6	86.0	80.7	88.2	0.81	-
	SSC-SleepNet (ours)	**Fpz-Cz**	**42308**	**93.6**	**60.2**	**92.4**	**91.6**	84.5	**89.0**	**84.5**	**90.1**	**0.85**	0.8 hrs
	DeepSleepNet**[21] [20]	Pz-Oz	41950	88.1	37.0	82.7	77.3	80.3	79.8	73.1	-	0.72	-
	SleepEEGNet* [21]	Pz-Oz	42308	90.3	44.6	85.7	81.5	**82.9**	82.8	77.0	-	0.77	-
	ResAtten* [82]	Pz-Oz	42308	87.2	36.8	85.2	81.3	80.1	80.7	74.1	-	0.74	-
	SSC-SleepNet (ours)	**Pz-Oz**	**42308**	**91.3**	**51.5**	**87.6**	**86.1**	82.8	**84.5**	**79.9**	**87.5**	**0.79**	**0.8 hrs**
Sleep-EDF-X	DeepSleepNet** [22] [20]	Fpz-Cz	195479	90.9	45.0	79.2	72.7	71.1	78.8	71.8	81.6	0.70	7.2 hrs**
	DeepSleepNet** [64] [20]	Fpz-Cz	195479	90.8	44.8	78.5	67.9	71.3	76.9	70.7	-	0.69	-
	DeepSleepNet-Lite* [81]	Fpz-Cz	163948	91.5	46.0	82.9	79.2	76.4	80.3	75.2	-	0.73	-
	AttnSleep* [22]	Fpz-Cz	195479	92.0	42.0	85.0	82.1	74.2	81.3	75.1	83.6	0.74	**1.7 hrs**
	U-Sleep* [83]	Majority vote	-	80.0	58.0	**88.0**	64.0	**91.0**	-	76.0	-	-	-
	SeqSleepNet** [22] [23]	Fpz-Cz	195479	91.8	46.0	85.0	77.5	81.0	82.6	76.3	84.3	0.76	7.3 hrs**
	XSleepNet1* [63]	Fpz-Cz, EOG	222479	92.6	50.2	85.9	79.2	81.3	83.6	77.8	-	0.77	-
	XSleepNet2* [63]	Fpz-Cz, EOG	222479	93.3	49.9	86.0	78.7	81.8	84.0	77.9	-	0.78	-
	SleepTransformer* [84]	Fpz-Cz	222479	91.7	40.4	84.3	77.9	77.2	81.4	74.3	-	0.74	-
	SleepEEGNet* [21]	Fpz-Cz	222479	91.7	44.1	82.5	73.5	76.1	80.0	73.6	-	0.72	-
	SleepEEGNet**[22] [21]	Fpz-Cz	195479	89.8	42.1	75.2	70.4	70.6	74.2	69.6	82.3	0.66	4.6 hrs**
	LGSleepNet* [64]	Fpz-Cz	195479	92.6	43.7	85.5	83.0	74.9	82.3	76.0	84.9	0.75	-
	SSC-SleepNet (ours)	**Fpz-Cz**	**195479**	**93.0**	**58.3**	86.4	**87.3**	76.8	**84.5**	**80.4**	**89.6**	**0.79**	1.8 hrs
	SleepEEGNet* [21]	Pz-Oz	222479	90.3	42.2	79.7	**94.9**	72.2	77.6	70.0	-	0.69	-
	SSC-SleepNet (ours)	**Pz-Oz**	**195479**	**90.8**	**52.7**	**85.9**	87.5	**73.2**	**82.5**	**78.0**	**86.4**	**0.76**	**1.8 hrs**
SHHS	DeepSleepNet** [22] [20]	C4-A1	324854	85.4	40.5	82.5	79.3	81.9	81.0	73.9	82.6	0.73	14.4 hrs**
	AttnSleep* [22]	C4-A1	324854	86.7	33.2	87.1	87.1	82.1	84.2	75.3	84.0	0.78	2.1 hrs
	SeqSleepNet** [22] [23]	C4-A1	324854	84.2	47.3	87.2	85.4	**88.6**	85.6	78.5	85.4	0.80	15.2 hrs**
	SleepEEGNet** [22] [21]	C4-A1	324854	81.3	34.4	73.4	75.9	77.0	73.9	68.4	82.7	0.65	6.4 hrs**
	U-Sleep* [83]	Majority vote#	-	93.0	51.0	87.0	76.0	92.0	-	80.0	-	-	-
	SeqSleepNet**# [80] [23]	C4-A1	1637525	91.8	49.1	88.2	83.5	88.2	87.2	80.2	-	0.82	-
	SeqSleepNet**# [8] [23]	C4-A1, EOG, EMG	1637525	91.4	43.3	87.4	82.9	87.3	86.5	78.5	-	0.81	-
	XSleepNet1*# [63]	C4-A1, EOG, EMG	1637525	91.6	51.4	88.5	85.0	88.4	87.6	80.7	-	0.83	-
	XSleepNet2*# [63]	C4-A1, EOG, EMG	1637525	92.0	49.9	88.3	85.0	88.2	87.5	81.0	-	0.83	-
	SleepTransformer*# [84]	C4-A1, EOG, EMG	1637525	92.2	46.1	88.3	85.2	**88.6**	87.7	80.1		0.83	-
	L-SeqSleepNet*# [80]	C4-A1	1637525	93.1	51.1	89.0	84.9	89.8	88.4	81.4	-	0.84	
	SSC-SleepNet (ours)	**C4-A1**	**324854**	**86.8**	**60.5**	**88.3**	**90.8**	81.5	**86.3**	**81.5**	**87.4**	**0.81**	**2.1 hrs**
	SSC-SleepNet# (ours)	**C4-A1**	**1637525**	**93.5**	**57.8**	**91.0**	**90.1**	87.8	**89.9**	**84.0**	**89.5**	**0.86**	**12.1 hrs**
HMC	**SSC-SleepNet (ours)**	**F4-M1**	**137243**	**83.4**	**55.2**	**82.5**	**87.6**	**74.9**	**79.3**	**76.7**	**85.4**	**0.73**	**1.9 hrs**
	SSC-SleepNet (ours)	**C4-M1**	**137243**	**70.5**	**55.0**	**81.2**	**85.9**	**73.7**	**76.7**	**73.3**	**83.1**	**0.69**	**1.9 hrs**
	SSC-SleepNet (ours)	**O2-M1**	**137243**	**78.0**	**53.1**	**80.2**	**83.2**	**70.0**	**75.4**	**72.8**	**82.9**	**0.68**	**1.9 hrs**
	SSC-SleepNet (ours)	**C3-M2**	**137243**	**78.6**	**51.8**	**82.1**	**86.0**	**74.9**	**77.7**	**74.7**	**84.1**	**0.71**	**1.9 hrs**

Note that, on all datasets, * represents the results reported in previous works, are not compatible for a direct comparison here due to the discrepancies in data split, the number of channel used, modelling tasks, etc. ** represents the results reported on second previous works, which may cause different results due to the different environments or parameters, etc., the first [ref.] represents the second report work and the second [ref.] represents the original work reports. #Majority vote in U-Sleep, that is, the hypnograms were generated using predictions from all available EEG-EOG channel combinations within each record. In SHHS dataset, different models cannot comparation directly due to the different data preprocessing or training split, # represents the results reported in previous works in the whole SHHS-5463 dataset (the number of subjects is 5463 and the experimental setup is train/test: 0.7/0,3), U-Sleep model uses 5767 subjects and the experimental setup is leave-one-out cross validation, and others use SHHS-329 datasets and the experimental setup is 20-fold cross validation.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3572886

population (a random mix of OSA and non-OSA patients), as shown in Table IX. Interestingly, the performance of non-OSA patients was significantly lower than in the general population, except for stage N1. Possible reasons for this are the following. First, the larger dataset size of the general population results in better or more stable training, favoring the classes with higher data representation. Second, the overnight sleep of OSA patients, characterized by frequent apnea events, often corresponds to interrupted sleep patterns (e.g. sleep fragmentation), which may allow the model to better identify these characteristics. We also validated the model on the HMC dataset, which includes patients with different sleep disorders. The results were clearly worse than those on the other datasets, indicating still the general challenge of sleep staging from single-channel EEG in patients with sleep disorders. Nevertheless, the N1 sleep detection performance (57.8% using the C4-M1 channel) on the HMC dataset were consistent with that on the other datasets.

E. Comparison with State-of-the-art Models

As presented in Table IX, SSC-SleepNet was able to effectively identify different sleep stages, with a higher accuracy, MF1 score, and Cohen's kappa compared with the results reported in literature. However, the result in detecting N3 or REM sleep was sometimes lower than that using other models. For example, the U-Sleep algorithm reported a higher REM sleep detection result than our work but at the expense of much lower result in NREM sleep detection. When using the Pz-Oz channel on the Sleep-EDF-X dataset, we obtained a worse detection of N3 sleep compared with that using SleepEEGNet while we did clearly better in detecting all the other stages. In terms of model performance, the use of a complex pseudo-Siamese learning network structure and repeated iterations of the adaptive loss function increased computational complexity. In particular, the adopted dynamic loss required multiple autocorrelation calculations on the input data, contributing to the higher computational requirements. As a result, the parameters and FLOPs of our model are relatively high, making it less suitable for portable and lightweight sleep staging applications. This limitation highlights an important direction for future work - the development of lightweight models that can maintain strong N1 detection performance while reducing computational complexity.

F. Limitations

There are several limitations in our study. First, we could not always directly compare model performance using the same channels in the four datasets due to different channels being available because of the use of different setups. However, we have demonstrated that the proposed algorithm on different EEG channels has led to effective sleep staging, indicating its robustness against those different EEG channels tested Second, we solely conducted comparison experiments between patients with an AHI less than 5 and the general population on the SHHS dataset. Comparing results in various patient groups would require further investigation [87]. In addition, in the HMC dataset containing patients with sleep disorders, only the electrode phase difference of the standard EEG channel was recorded, which may be the main reason why the corresponding sleep staging performance was markedly lower than that of the other datasets. Third, the hyperparameters of the dynamic loss function used in SSC-SleepNet were adjusted based on accuracy rather than the kappa score. In sleep analysis, the kappa score is preferred because it considers the imbalance of different sleep stages throughout the night. However, the accuracy can be calculated per sleep stage separately, whereas the kappa score always depends on the negative samples as well. Therefore, we chose to leverage the per-class accuracy, as it is more consistent with the adaptive contrastive loss function proposed in this study. Finally, SSC-SleepNet was specifically designed for signals collected by PSG devices, which may potentially limit its applicability to signals collected using, for example, portable or wearable devices, due to, for example, the presence of more motion artifacts. When employing a new, wearable EEG sensor instead of traditional PSG EEG channels, the robustness of our proposed algorithm against noise such as motion artifacts should be further verified. Additionally, while SSC-SleepNet demonstrated improved N1 prediction, its model complexity is relatively high, compared to lighter models such as DeepSleepNet-Lite and L-SeqSleepNet. Therefore, future research should focus on developing lightweight, noise-resistant models to facilitate its potential deployment for wearable or less-obtrusive EEG-based sleep staging. Third, the hyperparameters of the dynamic loss function used in SSC-SleepNet are adjusted based on accuracy rather than the kappa score. In sleep analysis, the kappa score is preferred because it considers the imbalance of different sleep stages throughout the night. However, the accuracy can be calculated per sleep stage separately, whereas the kappa score always depends on the negative samples as well. Therefore, we chose to leverage the per-class accuracy, as it is more consistent with the adaptive contrastive loss function proposed in this study. Finally, SSC-SleepNet is specifically designed for signals collected by PSG devices, which may limit its applicability to signals from wearable or non-contact devices due to differences in data collection methods and source characteristics. Additionally, while SSC-SleepNet demonstrates effective N1 prediction, its model complexity is relatively high, resulting in greater computational demands compared to lighter models such as DeepSleepNet-Lite and L-SeqSleepNet. Future research will focus on developing lightweight, noise-resistant models to facilitate practical daily sleep monitoring.

VI. Conclusion

The newly developed SSC-SleepNet model demonstrates a particular ability to identify informative features using single-channel EEG data across different temporal and spatial scales to achieve accurate classification of sleep stages. This deep learning solution outperforms most current models in automatic sleep staging, using single-channel EEG signals, and makes significant progress in detecting the N1 sleep stage, closely matching the accuracy of the PSG-based gold standard of manual scoring.

REFERENCES

[1] C. A. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carrillo, and W. C. Dement, "Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients," *Sleep Medicine*, vol. 2, no. 5, pp. 389-396, 2001.

[2] O. Yildirim, U. B. Baloglu, and U. R. Acharya, "A deep learning model for automated sleep stages classification using PSG signals," *International Journal of Environmental Research and Public Health*, vol. 16, no. 4, pp. 599, 2019.

[3] W. W. Tryon, "Issues of validity in actigraphic sleep assessment," *Sleep*, vol. 27, no. 1, pp. 158-165, 2004.

[4] Troester, Matthew M, Stuart F Quan, Richard B Berry, David T Plante, Alexandre R. Abreu, Mohammed Alzoubaidi, Anuja Bandyopadhyay, et al. "The AASM Manual for the Scoring of Sleep and Associated Events. 3rd ed," *American Academy of Sleep Medicine*, 2023.

[5] Motin, Mohammod Abdul, et al. "Sleep-wake classification using statistical features extracted from photoplethysmographic signals." *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019.

[6] Hori T, Sugita Y, Koga E, et al. Proposed supplements and amendments to 'A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects', the Rechtschaffen & Kales (1968) standard[J]. *Psychiatry & Clinical Neurosciences*, 2001, 55(3).

[7] M. Gorgoni, A. D'Atri, G. Lauri, P. M. Rossini, F. Ferlazzo, and L. De Gennaro, "Is sleep essential for neural plasticity in humans, and how does it affect motor and cognitive recovery?," *Neural Plasticity*, vol. 2013, 2013.

[8] Hossain J L, Shapiro C M. The prevalence, cost implications, and management of sleep disorders: an overview[J]. *Sleep and Breathing*, 2002, 6(02): 085-102.

[9] Flemons W W, Douglas N J, Kuna S T, et al. Access to diagnosis and treatment of patients with suspected sleep apnea[J]. *American Journal of Respiratory and Critical Care Medicine*, 2004, 169(6): 668-672.

[10] Epstein L J, Kristo D, Strollo Jr P J, et al. Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults[J]. *Journal of Clinical Sleep Medicine*, 2009, 5(3): 263-76.

[11] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "Automatic sleep stage classification using single-channel eeg: Learning sequential features with attention-based recurrent neural networks." pp. 1452-1455.

[12] H. Phan, and K. Mikkelsen, "Automatic sleep staging of EEG signals: recent development, challenges, and future directions," *Physiological Measurement*, vol. 43, no. 4, pp. 04TR01, 2022.

[13] Z. Mousavi, T. Y. Rezaii, S. Sheykhivand, A. Farzamnia, and S. Razavi, "Deep convolutional neural network for classification of sleep stages from single-channel EEG signals," *Journal of Neuroscience Methods*, vol. 324, pp. 108312, 2019.

[14] A. Malafeev, D. Laptev, S. Bauer, X. Omlin, A. Wierzbicka, A. Wichniak, W. Jernajczyk, R. Riener, J. Buhmann, and P. Achermann, "Automatic human sleep stage scoring using deep neural networks," *Frontiers in Neuroscience*, vol. 12, pp. 781, 2018.

[15] Michielli N, Acharya U R, Molinari F. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals[J]. *Computers in Biology and Medicine*, 2019, 106: 71-81.

[16] Toma T I, Choi S. An end-to-end convolutional recurrent neural network with multi-source data fusion for sleep stage classification[C]//*2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, 2023: 564-569.

[17] Efe E, Ozsen S. CoSleepNet: Automated sleep staging using a hybrid CNN-LSTM network on imbalanced EEG-EOG datasets[J]. *Biomedical Signal Processing and Control*, 2023, 80: 104299.

[18] Jung K, Kim M, Chung W. Epoch-level and sequence-level multi-head self-attention-based sleep stage classification[C]//*2023 11th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2023: 1-6.

[19] Lin S, Wang Y, Wang Z. An Automatic Multi-Head Self-Attention Sleep Staging Method Using Single-Lead Electrocardiogram Signals[J].

[20] Supratak A, Dong H, Wu C, et al. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2017, 25(11): 1998-2008.

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

[22] Eldele E, Chen Z, Liu C, et al. An attention-based deep learning approach for sleep stage classification with single-channel EEG[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2021, 29: 809-818.

[23] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "SeqSleepNet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 3, pp. 400-410, 2019.

[24] Liu Y, Li X, Yang L, et al. A CNN-transformer hybrid recognition approach for sEMG-based dynamic gesture prediction[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1-16.

[25] Li W, Xue L, Wang X, et al. ConvTransNet: A CNN–transformer network for change detection with multiscale global–local representations[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-15.

[26] Mienye I D, Swart T G, Obaido G. Recurrent neural networks: A comprehensive review of architectures, variants, and applications[J]. *Information*, 2024, 15(9): 517.

[27] Li Y, Zheng W, Wang L, et al. From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition[J]. *IEEE Transactions on Affective Computing*, 2019, 13(2): 568-578.

[28] Park J, Kim J, Jung S, et al. ECG-signal multi-classification model based on squeeze-and-excitation residual neural networks[J]. *Applied Sciences*, 2020, 10(18): 6495.

[29] Kumar P R, Bonthu K, Meghana B, et al. Multi-class Brain Tumor Classification and Segmentation using Hybrid Deep Learning Network Model[J]. *Scalable Computing: Practice and Experience*, 2023, 24(1): 69-80.

[30] Li W, Gao J. Automatic sleep staging by a hybrid model based on deep 1D-ResNet-SE and LSTM with single-channel raw EEG signals[J]. *PeerJ Computer Science*, 2023, 9: e1561.

[31] Xia C, Xiong C, Yu P. Pseudo siamese network for few-shot intent generation[C]//*Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2021: 2005-2009.

[32] Zhou W, Shen N, Zhou L, et al. PSEENet: A pseudo-siamese neural network incorporating electroencephalography and electrooculography characteristics for heterogeneous sleep staging[J]. *IEEE Journal of Biomedical and Health Informatics*, 2024.

[33] I. M. Colrain, "Sleep and the brain," Neuropsychology Review, vol. 21, pp. 1-4, 2011.

[34] Hyde M. The N1 response and its applications[J]. *Audiology and Neurotology*, 1997, 2(5): 281-307.

[35] Vogel E K, Luck S J. The visual N1 component as an index of a discrimination process[J]. *Psychophysiology*, 2000, 37(2): 190-203.

[36] D. Shrivastava, S. Jung, M. Saadat, R. Sirohi, and K. Crewson, "How to interpret the results of a sleep study," *Journal of Community Hospital Internal Medicine Perspectives*, vol. 4, no. 5, pp. 24983, 2014.

[37] C. Sun, C. Chen, J. Fan, W. Li, Y. Zhang, and W. Chen, "A hierarchical sequential neural network with feature fusion for sleep staging based on EOG and RR signals," *Journal of Neural Engineering*, vol. 16, no. 6, pp. 066020, 2019.

[38] J. Lu, C. Yan, J. Li, and C. Liu, "Sleep staging based on single-channel EEG and EOG with Tiny U-Net," *Computers in Biology and Medicine*, vol. 163, pp. 107127, 2023.

[39] Danker-Hopfe H, Anderer P, Zeitlhofer J, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard[J]. *Journal of sleep research*, 2009, 18(1): 74-84.

[40] Danker-Hopfe H, Kunz D, Gruber G, et al. Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders[J]. *Journal of sleep research*, 2004, 13(1): 63-69.

[41] Penzel T, Zhang X, Fietze I. Inter-scorer reliability between sleep centers can teach us what to improve in the scoring rules[J]. *Journal of Clinical Sleep Medicine*, 2013, 9(1): 89-91.

[42] Rosenberg R S, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring[J]. *Journal of Clinical Sleep Medicine*, 2013, 9(1): 81-87.

[43] Rosenberg, Richard S., and Steven Van Hout. "The American Academy of Sleep Medicine Inter-Scorer Reliability Program: Sleep Stage Scoring." *Journal of Clinical Sleep Medicine*, 09, no. 01 (January 15, 2013): 81–87.

[44] Lee, Yun Ji, Jae Yong Lee, Jae Hoon Cho, and Ji Ho Choi. "Interrater Reliability of Sleep Stage Scoring: A Meta-Analysis." *Journal of Clinical Sleep Medicine*, 18, no. 1 (January 2022): 193–202.

[45] K. Mikkelsen, and M. De Vos, "Personalizing deep learning models for automatic sleep staging," arXiv preprint arXiv:1801.02645, 2018.
[46] D. Hendrycks, and K. Gimpel, "Gaussian error linear units (gelus)," arXiv preprint arXiv:1606.08415, 2016.
[47] Targ S, Almeida D, Lyman K. Resnet in resnet: Generalizing residual architectures[J]. arXiv preprint arXiv:1603.08029, 2016.
[48] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," *PloS One*, vol. 14, no. 3, pp. e0214587, 2019.
[49] Zhou J, Jiang T, Li Z, et al. Deep Speaker Embedding Extraction with Channel-Wise Feature Responses and Additive Supervision Softmax Loss Function[C]//*Interspeech*. 2019: 2883-2887.
[50] Ho Y, Wookey S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling[J]. *IEEE Access*, 2019, 8: 4806-4813.
[51] Zhang Z, Sabuncu M. Generalized cross entropy loss for training deep neural networks with noisy labels[J]. *Advances in Neural Information Processing Systems*, 2018, 31.
[52] Ross T Y, Dollár G. Focal loss for dense object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2980-2988.
[53] Danielsson P E. Euclidean distance mapping[J]. *Computer Graphics and Image Processing*, 1980, 14(3): 227-248.
[54] Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 18661-18673.
[55] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.
[56] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185-1194, 2000.
[57] Kemp B, Zwinderman A, Tuk B, et al. Sleep-edf database expanded[J]. *Physionet Org*, 2018.
[58] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline, "The National Sleep Research Resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351-1358, 2018.
[59] S. F. Quan, B. V. Howard, C. Iber, J. P. Kiley, F. J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, J. Robbins, and J. M. Samet, "The sleep heart health study: design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077-1085, 1997.
[60] D. Alvarez-Estevez, and R. Rijsman, "Haaglanden medisch centrum sleep staging database (version 1.0. 1)," *PhysioNet*, 2021
[61] B. Kemp, "Model-based monitoring of human sleep stages," 1987.
[62] Wolpert, Edward A. "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects." *Archives of General Psychiatry*, 20.2 (1969): 246-247.
[63] Phan H, Chén O Y, Tran M C, et al. XSleepNet: Multi-view sequential model for automatic sleep staging[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(9): 5903-5915.
[64] Shen Q, Xin J, Liu X, et al. LGSleepNet: an automatic sleep staging model based on local and global representation learning[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023.
[65] Y. Bengio, and Y. Grandvalet, "No unbiased estimator of the variance of k-fold cross-validation," *Advances in Neural Information Processing Systems*, vol. 16, 2003.
[66] Sors A, Bonnet S, Mirek S, et al. A convolutional neural network for sleep stage scoring from raw single-channel EEG[J]. *Biomedical Signal Processing and Control*, 2018, 42: 107-114.
[67] Chen Y, Lv Y, Sun X, et al. ESSN: An Efficient Sleep Sequence Network for Automatic Sleep Staging[J]. *IEEE Journal of Biomedical and Health Informatics*, 2024.
[68] T. O. Kvålseth, "Note on Cohen's kappa," *Psychological Reports*, vol. 65, no. 1, pp. 223-226, 1989.
[69] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," arXiv preprint arXiv:1802.03426, 2018.
[70] D. Giavarina, "Understanding bland altman analysis," *Biochemia Medica*, vol. 25, no. 2, pp. 141-151, 2015
[71] Chollet F & others (2015) Keras. GitHub. Retrieved from https://github.com/fchollet/keras
[72] T. Roth, Y. Dauvilliers, D. Guinta, S. Alvarez-Horine, E. Dynin, and J. Black, "Effect of sodium oxybate on disrupted nighttime sleep in patients with narcolepsy," *Journal of Sleep Research*, vol. 26, no. 4, pp. 407-414, 2017.
[73] Hunger R. Floating point operations in matrix-vector calculus[M]. *Munich, Germany: Munich University of Technology, Inst. for Circuit Theory and Signal Processing*, 2005.
[74] T. Roth, Y. Dauvilliers, E. Mignot, J. Montplaisir, J. Paul, T. Swick, and P. Zee, "Disrupted nighttime sleep in narcolepsy," *Journal of Clinical Sleep Medicine*, vol. 9, no. 9, pp. 955-965, 2013.
[75] L. A. Moctezuma, T. Abe, and M. Molinas, "EEG-based 5-and 2-class CNN for sleep stage classification," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 3211-3216, 2023.
[76] M. Lee, H.-G. Kwak, H.-J. Kim, D.-O. Won, and S.-W. Lee, "SeriesSleepNet: an EEG time series model with partial data augmentation for automatic sleep stage scoring," *Frontiers in Physiology*, vol. 14, 2023.
[77] D. T. Pham, and R. Mouček, "Automatic Sleep Stage Classification by CNN-Transformer-LSTM using single-channel EEG signal." pp. 2559-2563.
[78] Z. Leng, M. Tan, C. Liu, E. Cubuk, X. Shi, S. Cheng, and D. Anguelov, "PolyLoss: A Polynomial Expansion Perspective of Classification Loss Functions. arXiv 2022," arXiv preprint arXiv:2204.12511.
[79] B. P. Lucey, J. S. Mcleland, C. D. Toedebusch, J. Boyd, J. C. Morris, E. C. Landsness, K. Yamada, and D. M. Holtzman, "Comparison of a single-channel EEG sleep study to polysomnography," *Journal of Sleep Research*, vol. 25, no. 6, pp. 625-635, 2016.
[80] Phan H, Lorenzen K P, Heremans E, et al. L-SeqSleepNet: Whole-cycle long sequence modelling for automatic sleep staging[J]. *IEEE Journal of Biomedical and Health Informatics*, 2023.
[81] Fiorillo L, Favaro P, Faraci F D. Deepsleepnet-lite: A simplified automatic sleep stage scoring model with uncertainty estimates[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2021, 29: 2076-2085.
[82] Qu, Wei, et al. "A residual based attention model for EEG based sleep staging." *IEEE Journal of Biomedical and Health Informatics*, 24.10 (2020): 2833-2843.
[83] Perslev M, Darkner S, Kempfner L, et al. U-Sleep: resilient high-frequency sleep staging[J]. *NPJ Digital Medicine*, 2021, 4(1): 72.
[84] Phan H, Mikkelsen K, Chén O Y, et al. Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification[J]. *IEEE Transactions on Biomedical Engineering*, 2022, 69(8): 2456-2467.
[85] Zhou W, Shen N, Zhou L, et al. PSEENet: A pseudo-siamese neural network incorporating electroencephalography and electrooculography characteristics for heterogeneous sleep staging[J]. *IEEE Journal of Biomedical and Health Informatics*, 2024
[86] Phan, Huy, et al. "Towards more accurate automatic sleep staging via deep transfer learning." *IEEE Transactions on Biomedical Engineering*, 68.6 (2020): 1787-1798.
[87] Lu C, Pathak S, Englebienne G, et al. Channel Contribution in Deep Learning Based Automatic Sleep Scoring—How Many Channels Do We Need?[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022, 31: 494-505.
[88] Cielo C M, Tapia I E. Diving deeper: rethinking AHI as the primary measure of OSA severity[J]. *Journal of Clinical Sleep Medicine*, 2019, 15(8): 1075-1076.