LEEGNet: Lightweight EEG Sleep Stage Classification Network with Knowledge Distillation

Boqian Wang[1]
School of AI and Advanced Computing
XJTLU Entrepreneur College (Taicang)
Xi'an Jiaotong-Liverpool University
Suzhou, China
Boqian.Wang19@student.xjtlu.edu.cn

Yulong Li[1]
School of AI and Advanced Computing
XJTLU Entrepreneur College (Taicang)
Xi'an Jiaotong-Liverpool University
Suzhou, China
Yulong.Li19@student.xjtlu.edu.cn

Chao Zhang
School of AI and Advanced Computing
XJTLU Entrepreneur College (Taicang)
Xi'an Jiaotong-Liverpool University
Suzhou, China
Chao.Zhang1902@student.xjtlu.edu.cn

Angelos Stefanidis
School of AI and Advanced Computing
XJTLU Entrepreneur College (Taicang)
Xi'an Jiaotong-Liverpool University
Suzhou, China
Angelos.Stefanidis@xjtlu.edu.cn

Mian Zhou*
School of AI and Advanced Computing
XJTLU Entrepreneur College (Taicang)
Xi'an Jiaotong-Liverpool University
Suzhou, China
Mianzhou.PT@xjtlu.edu.cn

Jionglong Su*
School of AI and Advanced Computing
XJTLU Entrepreneur College (Taicang)
Xi'an Jiaotong-Liverpool University
Suzhou, China
Jionglong.Su@xjtlu.edu.cn

2024 IEEE 5th International Conference on Pattern Recognition and Machine Learning (PRML) | 979-8-3503-5592-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/PRML62565.2024.10779817

***Abstract*—Sleep disorders have become increasingly common in recent years, seriously affecting people's quality of life and health. According to the Chinese Sociological Association in 2023, 38.3% of adults in China are diagnosed with sleep disorders. There are many causes of sleep disorders such as stress, *e.g.,* depression, and sleep breathing disorders. Therefore, treating sleep disorders require different approaches based on the different symptoms, but understanding the cause of sleep disorders requires hours of manual labeling by sleep specialists, which is inefficient in the clinical neurology. Although deep learning-based automated sleep stage classification models have achieved clinical accuracy, high complexity and computational resource requirements limit their widespread application. To address this, we propose Lightweight EEG Network (LEEGNet), which introduces knowledge distillation for the first time in the sleep stage classification task. This approach significantly keeps the parameters and complexity to a minimum while maintaining an acceptable level of accuracy, making the model more easily deployed in mobile devices, reducing the time required for manual annotation and improving the speed of diagnosis in clinical neurology. The number of parameters in LEEGNet is 0.9M, which is 2.47 times smaller as compared to the 2.4M of parameters in the state-of-the-art (SOTA) model SleePyCo. Furthermore, LEEGNet runs five times faster on a single-core CPU, and achieves an accuracy of 82.4%, which is only 2.2% lower than the SOTA model.**

Keywords—EEG, Sleep Stage Classification, Knowledge Distillation.

I. INTRODUCTION

Sleep plays an important role in human physiological and psychological health and is an essential part of human life. However, sleep disorders such as Sleep Breathing Disorder [1], Periodic Leg Movements [2], and Insomnia [3] can reduce the quality of sleep and lead to sleep deficiency. Long-term sleep deficiency can lead to complications such as cardiovascular disease, tumors, hypertension and heart disease, which can adversely affect the quality of life and long-term health of patients [4]. Therefore, early diagnosis of sleep disorders is crucial, and timely classification as well as intervention of these problems can reduce the cost of treatment and the risk of complications, which is of great importance in clinical neurology [5].

The American Academy of Sleep Medicine [6] divide human sleep into five stages, *i.e.*, Wakefulness (W stage, >13 Hz), Non-Rapid-Eye-Movement (NREM stage, which is further divided into three different stages, N1, N2, and N3, which correspond to the frequencies 4-7Hz, 11-16Hz, 0.5-2Hz, respectively), and Rapid-Eye-Movement (REM stage, corresponding to the frequency 2-6Hz). Sleep structure refers to the pattern of sleep pattern throughout the night, cycling into and through the different sleep stages, with each cycle typically lasting 90-120 minutes. The structure of each sleep stage can vary depending on age, sleep quality and individual differences [7]. Therefore, understanding the structure of sleep in clinical medicine is very important for the diagnosing and treating sleep disorders, which requires the doctor to accurately detect each cycle in the sleep stage [32]. Polysomnography (PSG) [8] is considered the gold standard for studying and evaluating the structure of sleep, which consists of bio-signals related to physical activity due to brain activity (electroencephalography, EEG), eye movement (electrooculography), heart rhythm (electrocardiogram), limb activity (electromyography), pulse and respiration. In particular, EEG directly reflects brain activity during sleep, capturing subtle changes during sleep and ensuring the

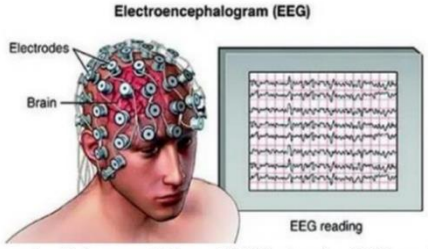

Fig. 1. Schematic of the acquisition of EEG signals, EEG reading is a common data format for EEGs [9].

1 Equal Contribution
* Corresponding author

Authorized licensed use limited to: Xi'an Jiaotong-Liverpool University. Downloaded on January 05,2025 at 14:24:42 UTC from IEEE Xplore. Restrictions apply.

accuracy of sleep stage classification [8]. Therefore, EEG is the main tool used to determine sleep structure.

The EEG signal collection technique and EEG data recordings are depicted in Fig. 1. By wearing a non-invasive capture cap, electrodes are used to measure the bioelectrical signal changes in the cerebral cortex at the corresponding moment, and record them in EEG format. EEG signals can be classified into five major rhythms based on their frequency and characteristics, which are mainly generated by the body's activities such as sleep. Therefore, the characteristics of the five major rhythms are mainly analyzed in sleep-related diagnoses. Each of these rhythms corresponds to the frequency and the stage of sleep that occurs in the rhythms as shown in the Table 1. In order to better understand the structure of sleep, the recorder usually connects the frequency intervals of the sleep stages horizontally in a temporal relationship, and the data format is shown in Fig. 2. Manual classification of the different sleep stages of EEG is a difficult process. It usually requires hours of sleep specialists' time to determine sleep stages [11]. Therefore, automatic identification of sleep stages is of great clinical importance.

Table 1. The five major rhythms of the EEG, the frequency intervals corresponding to each rhythm, and the frequencies at which the different phases are located

Rhythm	Frequency	Sleep Stage
Delta	0.2-3 Hz	N3(0.5-2Hz) REM(2-6Hz)
Theta	3 - 8 Hz	N1(4-7Hz) REM(2-6Hz)
Alpha	8 - 12 Hz	N2(11-16Hz)
Beta	12 - 27 Hz	N2(11-16Hz)
Gamma	>27 Hz	W (>13Hz)

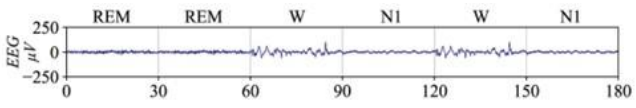

Fig. 2. Data image visualisation for a sleep stage classification.

The detection of brain electrical activity during sleep stages holds clinical significance. First, it aids in diagnosing and monitoring various sleep disorders such as insomnia, obstructive sleep apnea, and narcolepsy by identifying and analyzing sleep structure and quality, thus providing critical diagnostic information. Second, EEG can assess the impact of neurological diseases like epilepsy, Parkinson's disease, and Alzheimer's disease, helping to adjust treatment plans and understand disease progression. However, there are numerous challenges and limitations in the availability of medical resources to meet the needs related to sleep disorders:

1) Rural and remote areas often lack relevant medical resources, which makes it difficult to access effective diagnostic and treatment services for sleep disorders in these areas.

2) Diagnosis and treatment of sleep disorders are complex and diverse, requiring long-term, continuous monitoring and management, which imposes greater demands on the healthcare system.

3) High costs pose a significant barrier, as traditional monitoring equipment and treatments are expensive, making them unaffordable for many patients.

This research aims to address current limitations in the detection of sleep stage EEG signals within medical resources, enhancing healthcare service accessibility, improving diagnostic and treatment efficiency, as well as significantly reducing treatment costs and operational expenses for patients. We propose LEEGNet, a lightweight sleep stage classification network with knowledge distillation that address these limitations while guaranteeing its accuracy in classifying sleep stages. The key-contributions of our research are two-fold:

1) To the best of our knowledge, we are the first to introduce knowledge distillation for the purpose of lightweight modeling in sleep stage classification.

2) Our model reduces the number of parameters by a factor of 2.47 times and the CPU runtime by a factor of nearly five times compared to the SOTA model, achieving on accuracy of 82.4% that is only 2.2% lower than the SOTA model.

The rest of this paper is organized as follows: In Section 2 we introduce the relate work and its limitations, as well as how our model addresses them. In Section 3, we discuss in detail the design of our model. In Section 4, we present the training parameter for our model and all experimental results. Finally, in Section 5, the conclusion to our research is given, as well as the future research directions.

II. Related Work

Automatic classification of sleep stages based on EEG has been widely researched in recent years [12][13][14]. Early sleep stage classification tasks mainly use traditional machine learning algorithms such as Artificial Neural Network [15], Support Vector Machine (SVM) [19] and K-Nearest Neighbor (KNN) [20]. Many of the algorithms can only perform the sleep stage three classification task. They cannot classify the stages of NREM, *e.g.,* their results bear no clinical significance [16][17][18]. In the EEG-based classification task, Brignol *et al.* [19] propose a feature extraction method using time-frequency analysis combined with an SVM classifier to achieve 80% accuracy in predicting wakefulness stage, but the other categories' accuracies are very low. Mora *et al.* [20] propose using the wavelet transform for feature extraction and the KNN for classification to achieve 70% accuracy in this task. However, these machine learning algorithms are very computationally intensive, and due to the low classification accuracy, they are also difficult to apply in clinical medicine.

Deep learning algorithms generally achieve higher accuracy than traditional machine learning algorithms. As a

Authorized licensed use limited to: Xi'an Jiaotong-Liverpool University. Downloaded on January 05,2025 at 14:24:42 UTC from IEEE Xplore. Restrictions apply.

model which uses single-channel EEG as input, the U-Time [21] model uses a full Convolutional Neural Network (CNN) encoding-decoding structure to automatically learn effective features directly from the original EEG signal, and captures temporal features using jump connections. However, it requires a large amount of computation and does not fully capture temporal features. SleepEEGNet [22] combines CNN with Recurrent Neural Network (RNN) and demonstrates that the time series model has better performance in the EEG-based sleep stage classification task, but RNN suffer from performance degradation in long sequences. To address this, DeepSleepNet [23] combines RNN with Long Short-Term Memory (LSTM) to avoid the problem of RNN performance degradation when dealing with long sequences, but this requires a large number of computational resources, and model tuning becomes difficult. Phan *et al.* [24] introduce an attention mechanism to further improve the model's ability to capture long sequential dependencies and avoid the problem of difficult parameter tuning, but it still requires a large amount of data for training. In the model with multi-view EEG as input, mulEEG [25] uses contrast learning to improve the model's ability to extract correct features, which reduces the training time of the model and improves the robustness and generalization of the model, but the accuracy of this model is low. SleePyCo [26] uses CNNs and transformers as a model, combining comparative learning with a pyramid structure to make it a state-of-the-art (SOTA) model for this task by stacking convolutional layers and improving the model's ability to extract key features.

Current deep learning models suffer from high complexity and the need for large computational resources, both of which are difficult in applying these models in clinical medicine [27]. High complexity results in the models taking longer to obtain predicted results in the clinical medicine, and the model is less interpretability and transparency. High computational resource requirements make it difficult for models to inference on CPUs and the cost of using them increases, which can lead to many deployment bottlenecks, *e.g.*, high-cost consumption, high memory usage and poor scalability.

To address these, we propose a novel deep learning model LEEGNet, which greatly reduces the computational resource consumption of the model while ensuring high accuracy, providing technical possibilities for remote diagnosis and treatment and lightweight deployment of sleep stage classification in clinical medicine.

III. Methedology

The structure of our proposed LEEGNet model is shown in Fig. 3. It uses a knowledge distillation approach and consists of two main components: the teacher model and the student model. When training the student model, the teacher model being learnt needs to freeze the model parameters. The student model is a lightweight model we proposed, that requires to update the parameters so that the model approaches the feature extraction capability of the teacher model.

The overall model inputs L consecutive EEG epochs, each of which is acquired as single-channel 30s EEG epochs acquired at $s = 30$ samples/sec. The model classifies the input into different sleep stages for each EEG epoch. The input to the model is defined as $x \in R^{3000L \times 1}$, and the corresponding label is defined as $y = \{y_1, y_2, \ldots, y_L\}$ where $y_i \in \{0,1\}^{N_c}$, in which $i \in \{1,2, \ldots, L\}$ and N_c is the number of categories. In this training, the model has $N_c = 5$ categories, and $L = 10$.

A. Teacher Model

The teacher model structure is shown in Fig. 3. It uses SleePyCo [26], which consists of several modules: backbone network, lateral connections and classifier network. The model extracts multi-scale EEG feature sequences through a backbone network and forms a feature pyramid using side connections to capture features at different time and frequency scales. The classifier network performs in-depth analysis of these features through the transformer encoder and attention layers to accurately predict sleep stages during the target EEG period. To ensure that the student model in the knowledge distillation approach learns the complete

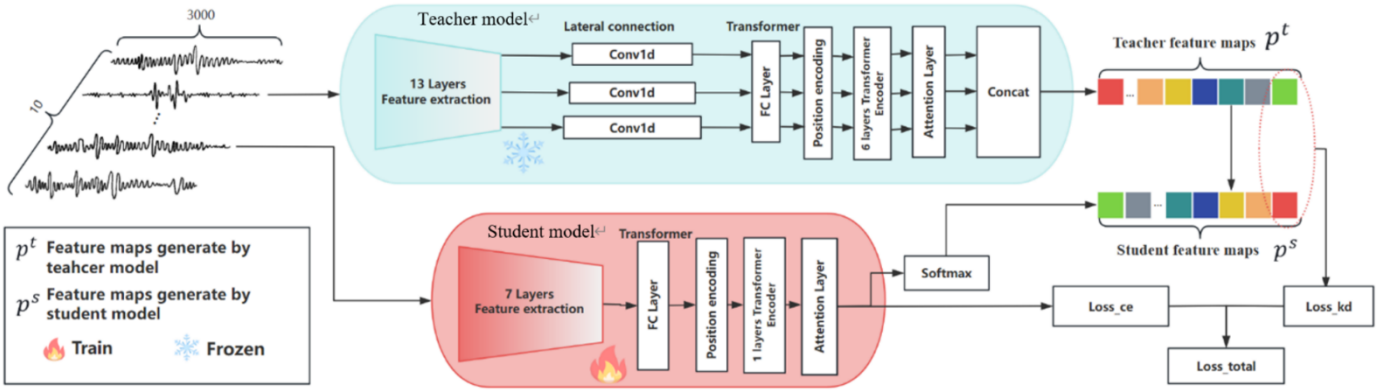

Fig. 3. Overall structure of LWEEGNet

Authorized licensed use limited to: Xi'an Jiaotong-Liverpool University. Downloaded on January 05,2025 at 14:24:42 UTC from IEEE Xplore. Restrictions apply.

feature maps of the teacher model, we carry out a splicing and merging of the feature maps of the teacher model into a single feature map. The teacher model will output three feature maps P^{t1}, P^{t2} and P^{t3} corresponding to different sizes of lateral connection, and the output of the final feature map p^t is calculated as:

$$p^t = \left(\frac{P_1^{t1} + P_1^{t2} + P_1^{t3}}{3}, \frac{P_2^{t1} + P_2^{t2} + P_2^{t3}}{3}, \dots, \frac{P_i^{t1} + P_i^{t2} + P_i^{t3}}{3}\right)$$

where $i = \{0, 1, \dots, N_f\}$ in which N_f is the length of feature map.

B. Student Model

The student model is given in Fig. 3. and the structure of the feature extraction module is shown in the Table 2. This part of the model greatly reduces the number of layers of the model, consisting of a transformer encoder using seven convolutional layers and one sampling layer.

Table 2. Detailed model structure of the feature extraction layer of the student model. Conv is 1-d convolution layer, BN is 1-d batch normalization, SE is the residual block.

Layer Name	Composition	Output Size
Input	-	30000 * 1
Conv1	3 Conv+BN+PReLU 3Conv +BN+SE+PReLU	30000 * 64
Max-pool1	5 Max-pool	6000 * 64
Conv2	3 Conv+BN+PReLU 3 Conv + BN + SE + PReLU	6000 * 128
Max-pool2	5 Max-pool	1200 * 128
Conv3	3 Conv+BN+PReLU 3 Conv+BN+PReLU 3 Conv+BN+SE+PReLU	1200 * 256
Max-pool3	5 Max-pool	240 * 256

C. Loss Function

The total loss $Loss_{total}$ of the model consists of the knowledge distillation loss $Loss_{kd}$ and the prediction loss $Loss_{ce}$, where the $Loss_{ce}$ is used to calculate the difference between the predicted results of the student model and the true value. The $Loss_{kd}$ is used to calculate the difference between the feature maps of the student model and the feature maps of the teacher model. The formulae for $Loss_{total}$ is :

$$Loss_{total} = w_1 Loss_{kd} + w_2 Loss_{ce}$$

where w_1 and w_2 are parameters for controlling the weight of loss.

Then, for the loss function of knowledge distillation $Loss_{kd}$, we use a hybrid loss function that combines soft knowledge distillation loss $Loss_{soft}$ and the hard knowledge distillation loss $Loss_{hard}$ to calculate $Loss_{kd}$. The formulae for $Loss_{kd}$ is:

$$Loss_{kd} = (1 - \alpha) Loss_{soft} + \alpha Loss_{hard}$$

where α is the parameter that control the loss weights.

We choose Mean Square Error as the $Loss_{hard}$ to make the feature maps of the student model close to the feature maps of the teacher model, so that the final result accuracy can approximate the teacher model. The formula for $Loss_{hard}$ is:

$$Loss_{hard} = \frac{\sum_{i=0}^{N_b} \sum_{j=0}^{N_f} (P_{i,j}^s - p_{i,j}^t)^2}{N_b N_f}$$

where N_b is the number of batch size, N_f is the length of feature map, $p_{i,j}^s = \{p_{i,1}^s, p_{i,2}^s, \dots, p_{i,N_c}^s\}$ are feature maps generated by student model where $i \in \{0,1\}^{N_f}$, $p_{i,j}^t = \{p_{i,1}^t, p_{i,2}^t, \dots, p_{i,N_c}^t\}$ are feature maps generated by teacher model where $i \in \{0,1\}^{N_f}$.

We choose Kullback-Leibler divergence [30] as the $Loss_{soft}$ to make the student model feature map close to the teacher model by comparing the probability distributions between the feature maps, preserving the ability of the student model to learn the data to improve the generalization ability of the model. The formula for $Loss_{soft}$ is:

$$Loss_{soft} = \sum_{i=0}^{N_b} \log\left(\frac{\exp(p_{i,j}^s)}{\sum_{j=0}^{N_f} \exp(p_{i,j}^s)}\right) \log\left(\frac{\log\left(\frac{\exp(p_{i,j}^s)}{\sum_{j=0}^{N_f} \exp(p_{i,j}^s)}\right)}{\frac{\exp(p_{i,j}^t)}{\sum_{j=0}^{N_f} \exp(p_{i,j}^t)}}\right)$$

The $Loss_{ce}$ is generated by calculating the Cross Entropy [31] between the predicted results of the student's model and the true value. Since Cross Entropy has good numerical stability, it is chosen as the loss function for classification. The formula for $Loss_{ce}$ is:

$$Loss_{ce} = -\log\left(\frac{\exp(x_j)}{\sum_j^{N_c} \exp(x_j)}\right) \quad (1)$$

where x_p is the output of the model, and N_c is the number of class.

IV. Experiment

A. Dataset

In this research, we use Sleep-EDF 2018 as our dataset, this dataset is the version published in 2018 [28], [29] and consists of 78 healthy Caucasian subjects aged 25-101 years. We divide the 78 subjects into 8 subjects of data for testing, 7 subjects for validation, and 63 subjects for training. We apply 10-fold cross-validation for our training. In this dataset, each subject records two full nights of PSG data, and some subjects collect one full night of PSG data. In this experiment, we use the single-channel 100 Hz Fpz-Cz EEG signal data from the PSG data, which is one-dimensional, and each 30s EEG signal epoch corresponds to a label.

Authorized licensed use limited to: Xi'an Jiaotong-Liverpool University. Downloaded on January 05,2025 at 14:24:42 UTC from IEEE Xplore. Restrictions apply.

B. Model Setting

The LEEGNet is trained using the Adam optimizer with a learning rate of 0.0005, a weight decay of 0.01 and a batch size of 128. The number of training rounds for the model is 100, due to the small size of the model in order to prevent overfitting, we use an early stopping strategy and set the patience of 7. This round of training is stopped when the training set accuracy is higher than the previous round's accuracy for a total of patience times. All the experiments are performed on a single-core run on an AMD 5950X CPU, where we freeze all the layers and set the batch size to 1 to facilitate the computation of the time used by the model predictions, avoiding the time error caused by batch scheduling.

C. Experiment Results

In this experiment, the metric used is Mean CPU, the average time required for a CPU to run. The results of our model are shown in Table 3, with the LWEEGNet model having only 0.9M parameters, which is 2.47 times fewer than that of SleePyCo. Moreover, the accuracy of our model on SleepEDF dataset is 82.4%, which is only 2.2% lower than the SOTA model. Our Mean CPU is nearly five times faster than SleePyCo. We trained with only student model, which has a classification accuracy of 74.1%, and the addition of knowledge distillation improves the accuracy by 7.7%, which proves that the use of knowledge distillation in LEEGNet is effective.

Table 3. **The best result is in bold.** The unit M in Parameter refers to millions. Comparison of the number of model parameters, runtime and accuracy, the best results is bolded. In the above, Parameter denotes the number of parameters of the model, and the unit M refers to millions. ACC refers to the 10-fold cross-validation accuracy of the model, and the unit is %. Mean CPU is the average CPU running time.

Method	Parameters(M)	ACC(%)	Mean CPU(ms)
DeepSleepNet [23]	2.1M	82.0	-
SleepEEGNet [22]	2.6M	80.0	-
mulEEG [25]	2.1M	79.6	-
SleePyCo [26]	2.4M	**84.6**	205.29
Student model	-	74.1	-
LEEGNet (Ours)	**0.97M**	82.4	**44.31**

D. Abliation Experiment

We carry out two sets of ablation experiments, one of which is a replacement of the transformer for the classification layer, while other is the modification of the feature extraction layer of the model. The purpose of these ablation experiments is to explore whether it is possible to further reduce the number of parameters in the model by reducing the parameters of the feature extraction and classification layers, while ensuring that the model's accuracy decreases are not excessively low.

For the first set of ablation experiments, we use GRU, bi-directional GRU, LSTM, and bi-directional LSTM for Transformer substitution, and the results and parameter sizes are given in Table 4. Both LSTM and GRU can significantly reduce the number of parameters compared to Transformer, but the decrease in accuracy is too obvious, for every 0.6M or so reduction in the number of parameters, the accuracy will decrease between 0.7% to 1%. The model change in the classification layer can reduce the number of parameters well, but when the accuracy of the model is reduced to about 80%, the model is difficult to be accepted in clinical applications.

Table 4. **The best result is in bold.** The unit M in Parameter refers to millions. Comparison of accuracy and parameters of classification layer modelling approach to substitution, the best result is bolded

Classifier Layer	Parameters(M)	ACC(%)
GRU	0.72M	79.6
Bidirection GRU	0.79M	80.1
LSTM	0.86M	80.8
Bidirection LSTM	0.91M	81.5
Transformer	**0.97M**	**82.4**

In the second set of ablation experiments, we observe the change in the accuracy of the model with respect to the number of parameters by further varying the number of feature extraction layers, we set up different layers to explore the effect of feature extraction layers on the model. The experimental results are given in Table 5. For every decrease in the number of feature extraction layers by two, the number of parameters in the model decreases by only 0.03M, and the accuracy decreases by 1.4%. From the results, reducing the feature extraction layer cannot significantly make the model more lightweight, and it may cause the accuracy of the model to decrease.

Tabel 5. **The best result is in bold.** The unit M in Parameter refers to millions. Modification of the number of convolutional layers in the feature extraction layer versus the resultant parameters and accuracies, the best result is bolded.

Method	Parameters(M)	ACC(%)
3-Layer	0.91M	79.4
5-Layer	0.94M	80.8
7-Layer	**0.97M**	**82.4**

V. Conclustion and Future Work

In this paper, we propose the LEEGNet architecture, a single-channel EEG sleep classification model in guaranteeing high accuracy. Leveraging on knowledge distillation techniques, our model uses SleePyCo as the teacher model, significantly reducing the number of parameters while maintaining high accuracy and improving the CPU operation speed by nearly five times. Compared to other deep learning models, LEEGNet's lightweight architecture allows it to run efficiently on devices with limited computational resources. In clinical medicine, LEEGNet can be used in portable devices, real-time monitoring, resource-limited environments, and telemedicine as a diagnostic aid for physicians to enhance the diagnosis and management of sleep disorders.

In spite of its efficacy, our model has three major limitations. First, we are currently unable to demonstrate that the model has good generalization capabilities. This means that we cannot determine how the model performs on different datasets, in particular whether it remains highly

Authorized licensed use limited to: Xi'an Jiaotong-Liverpool University. Downloaded on January 05,2025 at 14:24:42 UTC from IEEE Xplore. Restrictions apply.

accurate and robust across different domains or scenarios, which would result in the model only performing well on the dataset it is trained on. Second, the teacher model relies on pre-trained models from others, which limits the applicability of our model and may also introduce limitations inherent in external models. The original design of the teacher model and the training data may deviate from our task, thus affecting the effectiveness of our model. Third, although current model may perform well in experimental environments with sufficient computational resources, their computational complexity and storage requirements may become major bottlenecks in resource-constrained devices or real-time systems.

We identify three different research directions as future work. First, we will prioritize exposing our model to other datasets for ablation experiments to validate the generalization ability. Second, we will design a model with better performance for the sleep classification task as a teacher model, and use multi-task learning to improve the generalization ability of the teacher model before applying it to the lightweight student model. Third, we will reduce the complexity of the model by using low-complexity structures instead of fully connected layers, and use a model structure with high interpretability to improve model interpretability for better clinical deployment.

REFERENCES

[1] A. Malhotra and D. P. White, "Obstructive sleep apnoea," The Lancet, vol. 360, no. 9328, pp. 237–245, 2002.

[2] M. M. Ohayon and T. Roth, "Prevalence of restless legs syndrome and periodic limb movement disorder in the general population," Journal of Psychosomatic Research, vol. 53, no. 1, pp. 547–554, 2002.

[3] C. M. Morin and R. Benca, "Chronic insomnia," The Lancet, vol. 379, no. 9821, pp. 1129–1141, 2012.

[4] K. D. Kochanek, "Mortality in the United States, 2013," US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, No. 178, 2014.

[5] Y. Liu, L. Yan, B. Zeng, and W. Wang, "Automatic Sleep Stage Scoring Using Hilbert-Huang Transform with BP Neural Network," in 2010 4th International Conference on Bioinformatics and Biomedical Engineering, Chengdu, China, Jun. 2010. doi: 10.1109/icbbe.2010.5516372.

[6] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. F. Quan, The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, vol. 1. Westchester, IL: American Academy of Sleep Medicine, 2007.

[7] K. Guion and K. T. Avis, "Sleep architecture," in Encyclopedia of Child Behavior and Development, S. Goldstein and J. A. Naglieri, Eds. Boston, MA: Springer, 2011, doi: 10.1007/978-0-387-79061-9_2664.

[8] K. E. Bloch, "Polysomnography: a systematic review," Technology and Health Care, vol. 5, no. 4, pp. 285‑305, 1997.

[9] G. Wilson, "Rodents on my mind, rodents on the metro," NBB in Paris, Jun. 17, 2019. [Online]. Available: https://scholarblogs.emory.edu/nbbparis/2019/06/17/rodents-on-my-mind-rodents-on-the-metro/

[10] K. Majumdar, A Brief Survey of Quantitative EEG, 1st ed. CRC Press, 2017, doi: 10.1201/9781315117256.

[11] A. Malhotra *et al.*, "Performance of an automated polysomnography scoring system versus computer-assisted manual scoring," Sleep, pp. 573‑582, Apr. 2013, doi: 10.5665/sleep.2548.

[12] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "DNN filter bank improves 1-Max pooling CNN for single-channel EEG automatic sleep stage classification," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, doi: 10.1109/EMBC.2018.8512286.

[13] C. Sun, C. Chen, W. Li, J. Fan, and W. Chen, "A hierarchical neural network for sleep stage classification based on comprehensive feature learning and multi-flow sequence learning," IEEE Journal of Biomedical and Health Informatics, pp. 1351‑1366, May 2020, doi: 10.1109/J

[14] Y. Liu, L. Yan, B. Zeng, and W. Wang, "Automatic Sleep Stage Scoring Using Hilbert-Huang Transform with BP Neural Network," in 2010 4th International Conference on Bioinformatics and Biomedical Engineering, Chengdu, China, Jun. 2010. doi: 10.1109/icbbe.2010.5516372.

[15] T. Sousa, A. Cruz, S. Khalighi, G. Pires, and U. Nunes, "A two-step automatic sleep stage classification method with dubious range detection," Computers in Biology and Medicine, pp. 42‑53, Apr. 2015, doi: 10.1016/j.compbiomed.2015.01.017.

[16] S. Yu, X. Chen, B. Wang, and X. Wang, "Automatic sleep stage classification based on ECG and EEG features for daytime short nap evaluation," in Proceedings of the 10th World Congress on Intelligent Control and Automation, Beijing, China, 2012, pp. 4974-4977, doi: 10.1109/WCICA.2012.6359421.

[17] S. Gudmundsson, T. P. Runarsson, and S. Sigurdsson, "Automatic sleep staging using support vector machines with posterior probability estimates," in International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), Vienna, Austria, 2005, pp. 366-372, doi: 10.1109/CIMCA.2005.1631496.

[18] S. -F. Liang, C. -E. Kuo, Y. -H. Hu, and Y. -S. Cheng, "A rule-based automatic sleep staging method," in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 2011, pp. 6067-6070, doi: 10.1109/IEMBS.2011.6091499.

[19] A. Brignol, T. Al-ani, and X. Drouot, "EEG-based automatic sleep-wake classification in humans using short and standard epoch lengths," in 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 2012, pp. 276-281, doi: 10.1109/BIBE.2012.6399688.

[20] A. M. Mora *et al.*, "Sleeping with ants, SVMs, multilayer perceptrons and SOMs," in 2010 10th International Conference on Intelligent Systems Design and Applications, Cairo, Egypt, 2010, pp. 126-131, doi: 10.1109/ISDA.2010.5687278.

[21] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," Advances in Neural Information Processing Systems, vol. 32, 2019.

[22] S. Mousavi, F. Afghah, and U. R. Acharya, "SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach," PLoS ONE, vol. 14, 2019.

[23] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 11, pp. 1998-2008, Nov. 2017, doi: 10.1109/TNSRE.2017.2721116.

[24] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. D. Vos, "Automatic sleep stage classification using single-channel EEG: learning sequential features with attention-based recurrent neural networks," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, pp. 1452-1455, doi: 10.1109/EMBC.2018.8512480.

[25] V. Kumar, L. Reddy, S. Kumar Sharma *et al.*, "mulEEG: a multi-view representation learning on EEG signals," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022, pp. 398-407.

[26] S. Lee, Y. Yu, S. Back, H. Seo, and K. Lee, "Sleepyco: Automatic sleep scoring with feature pyramid and contrastive learning," Expert Systems with Applications, vol. 240, p. 122551, Apr. 2024, doi: 10.1016/j.eswa.2024.122551.

Authorized licensed use limited to: Xi'an Jiaotong-Liverpool University. Downloaded on January 05,2025 at 14:24:42 UTC from IEEE Xplore. Restrictions apply.

[27] G. Wilson, "Rodents on my mind, rodents on the metro," NBB in Paris, Jun. 17, 2019. [Online]. Available: https://scholarblogs.emory.edu/nbbparis/2019/06/17/rodents-on-my-mind-rodents-on-the-metro/

[28] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "DNN filter bank improves 1-Max pooling CNN for single-channel EEG automatic sleep stage classification," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, doi: 10.1109/EMBC.2018.8512286.

[29] C. Sun, C. Chen, W. Li, J. Fan, and W. Chen, "A hierarchical neural network for sleep stage classification based on comprehensive feature learning and multi-flow sequence learning," IEEE Journal of Biomedical and Health Informatics, pp. 1351‑1366, May 2020, doi: 10.1109/J

[30] [1] J. M. Joyce. (2011). Kullback-leibler divergence. Springer Berlin Heidelberg. DOI:10.1007/978-3-642-04898-2_327.

[31] A. Mao, M. Mohri, Y. Zhong. Cross-entropy loss functions: Theoretical analysis and applications. InInternational conference on Machine learning 2023 Jul 3 (pp. 23803-23828). PMLR.

[32] K. Guion, K.T. Avis. (2011). Sleep Architecture. In: Goldstein, S., Naglieri, J.A. (eds) Encyclopedia of Child Behavior and Development. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-79061-9_2664

Authorized licensed use limited to: Xi'an Jiaotong-Liverpool University. Downloaded on January 05,2025 at 14:24:42 UTC from IEEE Xplore. Restrictions apply.