DOI: https://doi.org/10.24297/ijct.v24i.9567

Convolutional Neural Networks for Deep Sleep Detection Based on Data Augmentation

Ruixuan Chen[1], *Linfeng Sui*[1,2], *Mo Xia*[1], *Jinsha Liu*[1], *Tao Zhang*[3] *and Jianting Cao*[1,2]

[1]Graduate School of Engineering, Saitama Institute of Technology, Fukaya 369-0217, Japan
[2]RIKEN Center for Advanced Intelligence Project (AIP), 103-0027, Japan
[3]College of Life Sciences, Nankai University, Tianjin, China

Email addresses {i3002hgt, i1002jhr, i2001fwi, f2004mdi}@sit.ac.jp; zhangtao@nankai.edu.cn;
Correspondence: cao@sit.ac.jp

Abstract

Sleep is a necessary process that individuals undergo daily for physical recovery, and the proportion of deep sleep in the sleep stages is a critical aspect of the recovery process. Convolutional Neural Networks (CNNs) have shown remarkable success in automatically identifying deep sleep stages through the analysis of electroencephalogram (EEG) signals. This article introduces three data augmentation techniques, including time shifting, amplitude scaling and noise addition, to enhance the diversity and features of the data. These techniques aim to enable machine learning models to extract features from various aspects of sleep EEG data, thus improving the model's accuracy. Three deep learning models are introduced, namely DeepConvNet, ShallowConvNet and EEGNet, for the identification of deep sleep. To evaluate the proposed methods, the Sleep-EDF public dataset was utilized. Experimental results demonstrate that the enhanced dataset formed by applying the three data augmentation techniques achieved higher accuracy in all deep learning models compared to the original dataset. This highlights the feasibility and effectiveness of these methods in deep sleep detection.

Keywords: deep sleep, data augmentation, convolutional neural networks, k-fold cross-validation

Introduction

Sleep is a key element in maintaining human health, with its quality having a profound impact on our overall well-being (Cirelli & Tononi, 2008). As modern life accelerates, lifestyle changes significantly disrupt regular sleep patterns, leading to various sleep disorders that severely affect societal safety and economic development (Ohayon, 2002). Adequate and high-quality sleep is crucial for strengthening the immune system and disease resistance. It also plays a significant role in enhancing memory, improving attention and boosting decision-making abilities (Barnes et al., 2015).

Electroencephalography (EEG) is a non-invasive method for tracking and studying the electrical activity in the brain (Niedermeyer & da Silva, 2005). This process involves placing electrodes on the scalp to measure and record the electrical signals emitted by neurons in the EEG. Compared to other methods such as functional magnetic resonance imaging (fMRI), computed tomography (CT), and positron emission tomography (PET), EEG offers advantages in terms of high temporal and spatial resolution, non-invasiveness, real-time monitoring and a wider range of applications (Michel & Murray, 2012). Therefore, EEG has become the most commonly used clinical diagnostic method for brain disorders, involving the recording of electrical activity to analyze brain function, and it is also utilized to aid in the diagnosis of sleep disorders (Li & Cao, 2023).

The classification and understanding of sleep stages rely heavily on the analysis of EEG frequency bands. Sleep is generally composed of two main stages: Rapid Eye Movement (REM) and Non-Rapid Eye Movement (NREM). According to the definition by the American Academy of Sleep Medicine (AASM) (Iber, 2007), NREM sleep is divided into three stages: Stage 1 (N1), associated with the θ wave band (4-8Hz), marks the onset of light sleep, characterized by a slowing down of muscle activity and increased ease of waking (Carskadon, Dement, et al., 2005); Stage 2 (N2), characterized by sleep spindles (12-15 Hz) and K-complexes, indicating progressively deeper sleep (De Gennaro & Ferrara, 2003); Stage 3 (N3), also known as Slow Wave Sleep (SWS), is dominated by δ waves (0-4Hz), representing deep sleep, a crucial period for tissue repair and memory consolidation (Stickgold, 2005). The REM sleep stage is characterized by low-amplitude, high-frequency mixed-frequency waves, similar to the β band (13-30Hz) and Gamma band (above 30Hz), during which brain activity resembles that of wakefulness, accompanied by rapid eye movements and dreaming (Hobson & Pace-Schott, 2002). The examples of EEG signals with a 30-second epoch of each sleep stage are shown in Fig. 1.

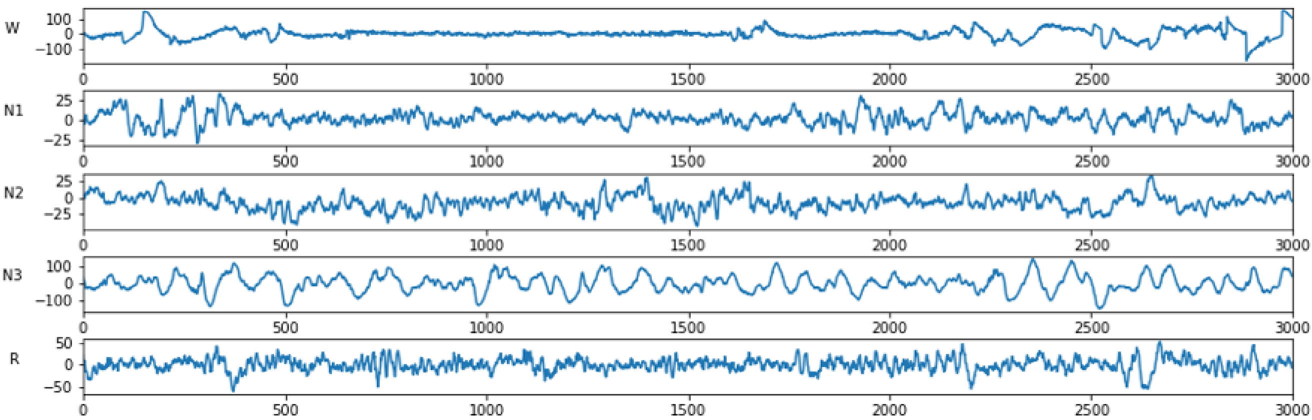

Fig. 1. Examples of EEG Data with a 30 seconds epoch of each sleep stages including Wake, N1, N2, N3 and REM.

Deep sleep, also known as Slow Wave Sleep (SWS), primarily occurs during the third stage of NREM sleep, accounting for about 20-25% of total sleep time in adults ((JSSR): et al., 2001). It represents a sleep cycle where both the body and brain enter a state of profound rest. During this stage, the brain generates Delta waves with frequencies in the 0-4Hz range, which take a dominant role in brainwave activity (Xia et al., 2023). These low-frequency, high-amplitude waveforms are the characteristic indicators of deep sleep. During deep sleep, there is a significant reduction in physiological activity: the heart rate and breathing pace slow down by about 20-30%, blood pressure drops by approximately 10-20%, and muscle tension and metabolic activity also decrease considerably. This sleep state is extremely important for the human body, as it not only aids in physical recovery and repair but is also crucial for memory consolidation and the healthy development of the brain (Caporale et al., 2021). A complete sleep cycle, which includes alternating stages of NREM and REM sleep, typically lasts between 70 to 110 minutes. In a typical good night's sleep, this cycle may repeat 4 to 6 times. These cycles follow a specific sequence, beginning with N1 (the onset of sleep), followed by N2 (light sleep) and N3 (deep sleep). After completing deep sleep, the cycle moves back to N2 before entering the REM phase. The transition from deep sleep (N3) back to light sleep (N1 or N2) leads into the REM sleep phase. This continuous process ensures the quality and efficiency of sleep, which is crucial for physical recovery and brain function (Basics, 2021).

Therefore, as long as professionals can detect the distribution and proportion of deep sleep during a person's sleep, they can effectively assess sleep quality. However, manually analyzing sleep EEG data in clinical diagnosis is an extremely time-consuming and energy-intensive task, reliant on highly skilled and experienced doctors. At the same time, EEG signals are very weak and unstable with strong chaotic characteristics, and they are accompanied by significant background noise. Traditionally, the interpretation of EEG has been manual. Due to the labor intensity and the subjectivity involved, it's difficult to implement EEG tests on a large scale for screening purposes. Monitoring a patient's sleep can require an extended period, sometimes up to a month of recording their EEG data. Such prolonged monitoring may cause psychological stress for the patient and also implies that doctors have to manage a considerable amount of data. Therefore, this method is challenging to disseminate widely among the general population (Chen et al., 2008).

In this paper, we propose three data augmentation methods for feature extraction in sleep EEG data: Time Shifting (TS), Amplitude Scaling (AS),and Noise Addition (NA). These methods are used to enhance the analysis of sleep EEG data by applying certain transformations to existing data to generate more samples (Shorten & Khoshgoftaar, 2019). The advantage of this approach is that it helps the model learn data features better and improves the model's generalization ability. This not only shortens the period for patients to record sleep but also reduces the workload of doctors to some extent. In the classification and grading stage, we propose three deep learning models: the DeepConvNet model (Mekruksavanich & Jitpattanakul, 2022), the ShallowConvNet model and the EEGNet model (Lawhern et al., 2018; Roots et al., 2020), which have become standard tools in deep learning research. DeepConvNet, a model based on convolutional neural networks, excels in handling multi-channel EEG data and accurately determining different sleep stages. In contrast, ShallowConvNet, a more streamlined convolutional neural network (CNN), is optimized for time-domain feature processing, characterized by fewer convolutional and pooling layers, reduced parameters and quicker training times. EEGNet merges CNN with Temporal Convolutional Network (TCN) in a hybrid format, adept at leveraging the spatiotemporal characteristics of EEG data. This model features a compact structure and localized connectivity, adept at capturing both temporal and frequency domain details in EEG signals (Zhao et al., 2023). To assess the effectiveness of our data augmentation strategies in enhancing network model training and feature extraction, we utilized the Sleep-EDF public datasets. This approach helped confirm the practicality and efficiency of using data augmentation methods for feature extraction in sleep stage detection.

The rest of this paper is organized as follows: Section 2 describes the data augmentation methods used. Section 3 introduces the architectures of the three deep learning models. In Section 4, we introduce a validation method involving K-fold cross-validation. Section 5 presents the database, device and experiments, followed by the experimental results and the conclusion of the paper.

Data Augmentation

Artificial data augmentation holds key importance in machine learning, especially when dealing with biological data that often shows notable rarity and heterogeneity (Anicet Zanini & Luna Colombini, 2020). Employing data augmentation is a prevalent approach in extracting features from signals. This process entails isolating valuable and distinct features from the collected data, which aids in enhancing learning, broadening generalization and easing human comprehension (Sarangi et al., 2020). Such methods are instrumental in boosting the efficacy of models tasked with classifying sleep stages. The features thus extracted can be utilized as inputs for various classification algorithms in machine learning, including Support Vector Machines (SVM) (Cortes & Vapnik, 1995), Artificial Neural Networks (ANN) or Random Forests (RF) (Guresen & Kayakutlu, 2011; Ho et al., 1995). This facilitates precise and dependable tracking of cerebral activity during sleep. In our research, we aim to minimize the data volume needed for accurate classification, consequently reducing the data-gathering burden on participants. Through a range of transformations and syntheses applied to the original signal, data augmentation generates augmented datasets, thereby expanding data variety, enhancing the resilience of models, addressing issues of data imbalance, and fostering the extraction of more comprehensive features (Shorten & Khoshgoftaar, 2019). The following are examples of augmented data application in signal feature extraction:

Time Shifting. By shifting the original EEG data along the time axis, especially when dealing with sleep data, we create a new time-shifted signal. This process involves moving the EEG data collected during sleep forward or backward by a certain number of time steps, effectively simulating the temporal variation of brain activity during sleep. The altered data thus presents a time-delayed version of the original signal. This technique helps the model recognize temporal correlations within sleep data and enhances its understanding of the EEG data features' changes during different sleep stages, such as the variations in frequency rhythms (α waves, β waves, etc.) across light sleep, deep sleep and REM sleep. Time shifting offers an in-depth understanding of the temporal dynamics in sleep data. In the context of sleep EEG data, this time shifting helps the model accurately capture the relationships and dynamic changes between deep sleep and other various sleep stages (Iwana & Uchida, 2021). Implementing time shifting not only enables the model to better adapt to minor variations and shifts in sleep data but also improves the model's overall robustness in detecting sleep stages. Fig. 2 shows an example of sleep EEG data after undergoing 1000 instances of time shifting.

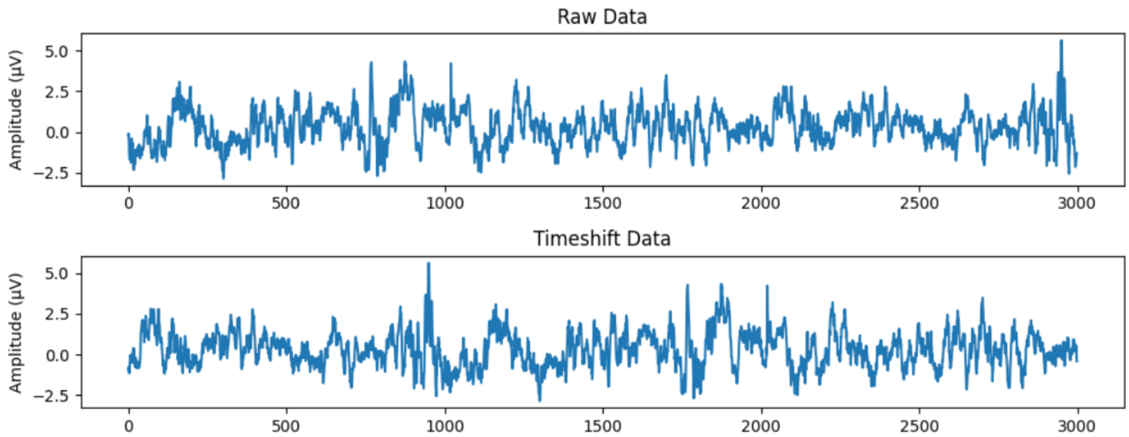

Fig. 2. RAM EEG Data and TS EEG Data

Amplitude Scaling. Amplitude scaling applied to EEG data can emphasize key features within the data. This method effectively highlights the signal patterns and variations between deep sleep and other sleep stages, making it easier for the model to identify key information. Transforming the Raw EEG data into a lower-dimensional feature representation can simplify the feature space, potentially enhancing system efficiency and performance. Adjusting the amplitude range of EEG data helps minimize the impact of noise, thus improving the accuracy and stability of the model (Sanei & Chambers, 2013). Amplitude scaling ensures consistency of data across various amplitude ranges, enabling effective comparison and analysis of data from diverse sources, sampling rates, or EEG equipment. This is vital for building a universal model capable of classifying different sleep stages. Assuming the original EEG data is $x(t)$, the amplitude-scaled signal can be represented as $ax(t)$, where a is the scaling factor, either a value greater than 1 for amplification or less than 1 for reduction. Fig. 3 shows an example of sleep EEG data that has undergone amplitude scaling.

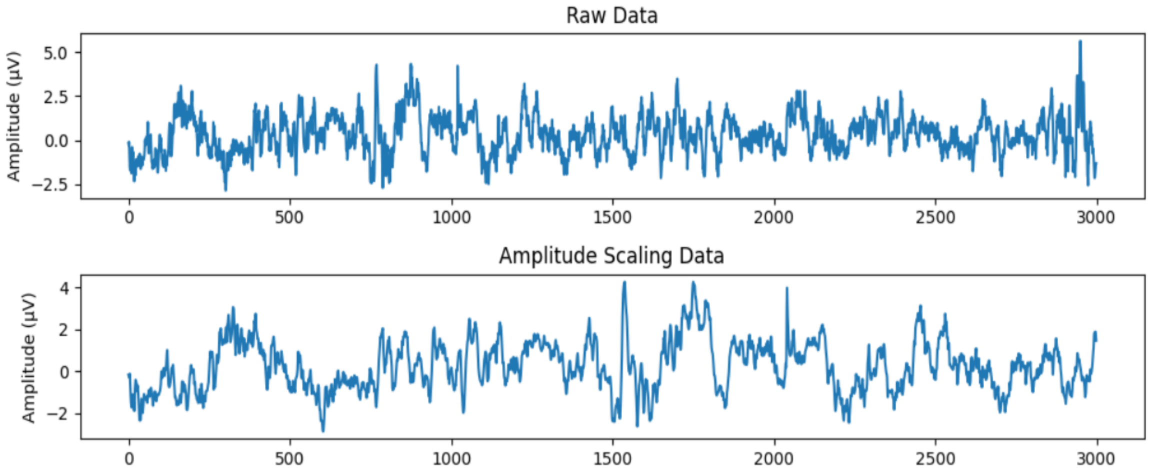

Fig. 3. RAM EEG Data and AS EEG Data

Noise Addition. Noise addition is a commonly used data augmentation technique in sleep data processing, enhancing data diversity by introducing randomness and variation. In the analysis of sleep EEG data, adding various types and intensities of noise to the Raw data simulates a wide range of data variations and noise interferences that may occur during sleep, thereby enabling the model to better adapt and process various sleep data. This method is particularly helpful in improving the robustness and generalization of the model when dealing with small or imbalanced sleep datasets. By introducing moderate levels of noise, the model becomes more adaptable to real-world variations and noise conditions found in sleep data. In this study, we employed the Gaussian noise method to enhance sleep EEG data. The Gaussian noise method involves generating random numbers that follow a Gaussian distribution with a mean of 0 and a specified variance, and adding these numbers to each feature of the sleep EEG data. Fig. 4 shows an example of sleep EEG data after noise addition.

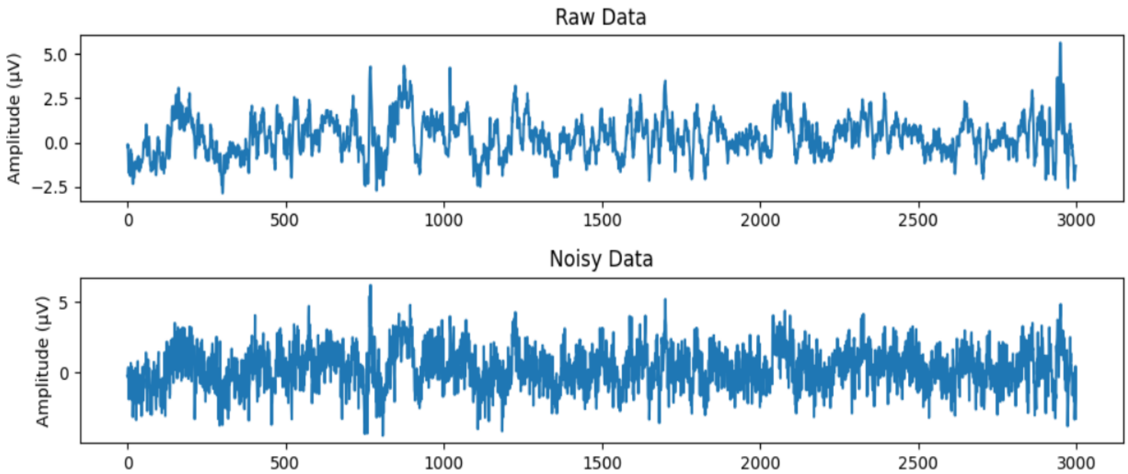

Fig. 4. RAM EEG Data and NA EEG data

Convolutional Neural Networks

In this paper, we introduce three deep learning models: DeepConvNet, ShallowConvNet and EEGNet, all of which employ CNN architectures, focusing on enhancing the accuracy and efficiency of sleep stage classification. DeepConvNet is a 1D-CNN-based model designed specifically for processing one-dimensional sequential data like sleep EEG data (Eren et al., 2019). This model is particularly suitable for analyzing and identifying deep sleep and other sleep stages in sleep EEG data, as it is capable of effectively capturing subtle changes in time-series data. The ShallowConvNet model, with fewer parameters and a shallow architecture, is ideal for simpler classification tasks, such as basic sleep stage classification (like wakefulness, light sleep, deep sleep) (Roots et al., 2020). Its simplicity also results in faster training and inference speeds, crucial for quickly analyzing large volumes of sleep data. EEGNet, designed specifically for EEG data, takes into account both temporal and spatial correlations, as well as spectral features (Lawhern et al., 2018). This provides it with unique advantages in automatically detecting and classifying sleep stages. Its lightweight architecture, robustness, and interpretability make it an efficient tool, especially in handling complex sleep patterns in EEG data. In summary, these three models each have their unique features and collectively offer a comprehensive and effective solution for the analysis of sleep EEG data, particularly in the automatic classification and recognition of sleep stages.

DeepConvNet Model. The DeepConvNet model utilizes a one-dimensional convolutional neural network (1D-CNN) to directly classify raw EEG signals without the need for any feature extraction stage, as the 1D-convolutional layer is capable of extracting features from EEG data. The developed DeepConvNet model incorporates five different types of layers: Convolutional layer, Dropout layer, Batch Normalization (BN) layer, Pooling layer and Fully Connected (FC) layer. The architecture of the DeepConvNet model is illustrated in Fig. 5.

In each layer, convolutional kernels of size 5×1 are used to reduce the computational load during the feature extraction stage. To ensure no features are missed, the stride of the convolutional kernel is set to 1. All convolutional layers utilize the Rectified Linear Unit (Relu) activation function. To reduce the computational burden of the entire model while not missing critical features, max-pooling layers with a pool size of 2×1 and a stride of 2 are set after each convolutional layer. All feature maps obtained from the Flatten layer are flattened into a one-dimensional feature vector, serving as the input for the first Fully Connected layer. The output from the first fully connected layer is nonlinearized using the Relu activation function and dropout with a rate of 0.5, and then serves as the input for the second Fully Connected layer. In this layer, the input EEG signals are classified into deep sleep and other sleep stages. To reduce the effect of overfitting, Dropout layers and Batch Normalization layers are added at various locations. The Batch Normalization layer normalizes the output from the upper max-pooling layer.

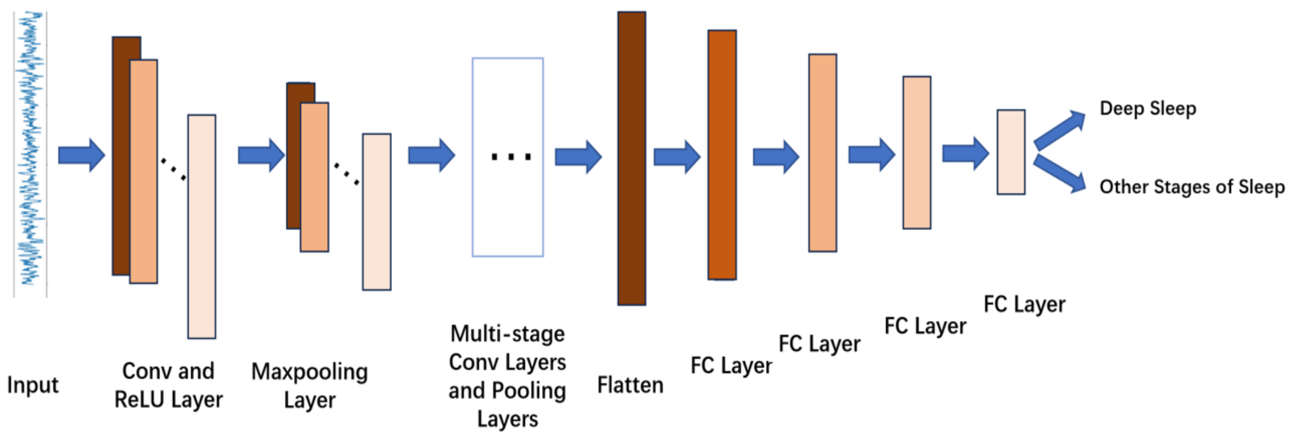

Fig. 5. Structure of DeepConvNet model

ShallowConvNet Model. ShallowConvNet, a specialized CNN design, is optimized for EEG data analysis. Tailored for EEG data classification, it's particularly useful in brain-computer interface (BCI) applications and the evaluation of cognitive states.

This network is structured into two primary sections. The initial section incorporates a sequence of convolutional layers that process the EEG input to extract both spatial and temporal characteristics. This section comprises a 1D convolutional layer, augmented by batch normalization, ReLU activation and average pooling, adept at identifying local patterns and spatial connections within the EEG data. The subsequent section of ShallowConvNet transforms the output from the first block into a flattened format, linking it to a dense (fully connected) layer. This dense layer correlates directly with the intended classification category and is succeeded by a softmax activation function that generates a class-based probability distribution. The design of the ShallowConvNet is depicted in Fig. 6.

Notable for its streamlined architecture, ShallowConvNet has fewer layers and parameters compared to more complex CNN models, enhancing its computational efficiency and making it ideal for real-time operations. The network is engineered to discern both spatial and temporal relationships in EEG data, enabling precise differentiation of various brain states or classifications (Roots et al., 2020). It is also trainable with limited datasets, making it versatile in settings with constrained data availability. In essence, ShallowConvNet stands out as a practical, lightweight model for EEG data classification, widely recognized for its simplicity and efficacy in both neuroscientific research and clinical contexts.

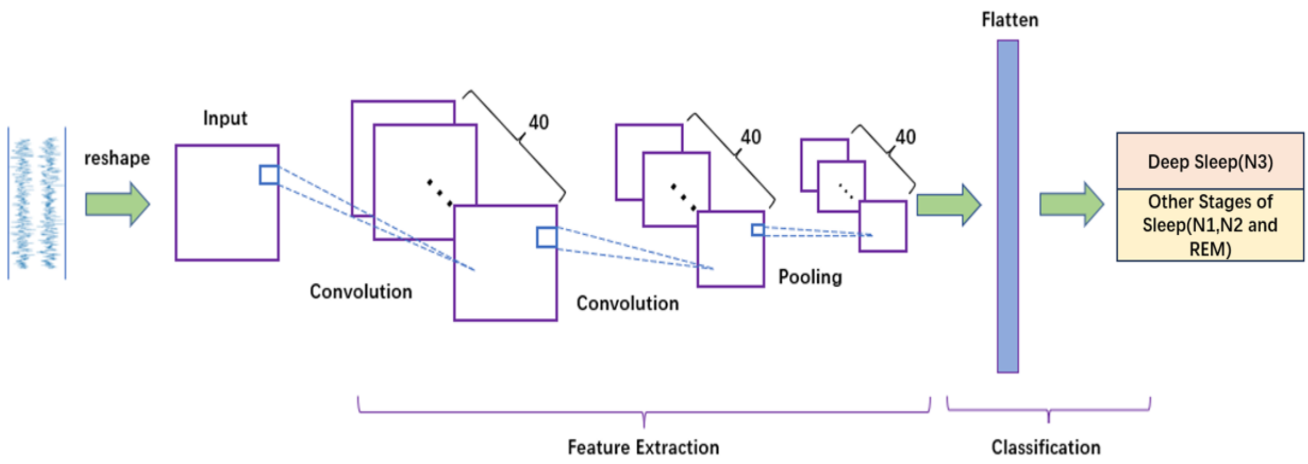

Fig. 6. Structure of ShallowConvNet Model

EEGNet Model. EEGNet is a deep learning model specifically designed for EEG signals (Lawhern et al., 2018). Due to its relatively complex network structure, while limiting the number of parameters, and its outstanding performance in various EEG-based tasks with low computational demands, EEGNet has seen widespread application. The EEGNet model includes Block 1, Block 2 and a classification block, with its structure illustrated in Fig. 7.

In Block 1, two convolutional steps are performed sequentially. The raw EEG data is reshaped into an input of size 1×50, initially convolved by a 1×64 filter with a stride of 8, followed by a Batch Normalization (BN) layer. Subsequently, each pooling layer uses a 1×4 average kernel size, with BN applied. Before applying the Exponential Linear Unit (ELU), batch normalization is applied along the feature map dimension. To help regularize the model, the dropout technique is utilized and the dropout is set for within-subject classification to prevent overfitting.

In Block 2, we employ 16 separable convolutions with a kernel size of 1×16, which is a depth convolution layer. The main role of separable convolutions is to reduce the number of parameters in the feature map, then optimally merge the outputs, explicitly decoupling the relationships within and across feature maps. After this, batch normalization, ELU and dropout are applied, similar to Block 1. Finally, an average pooling layer with a kernel size of 1×8 is used for dimension reduction.

In the classification part, the extracted features are transformed into a one-dimensional format and divided into four different categories using the SoftMax function. As SoftMax is a smooth and differentiable activation function, it enables efficient computation of its gradients, which is suitable for backpropagation and weight updates during the training process (Bishop & Nasrabadi, 2006).

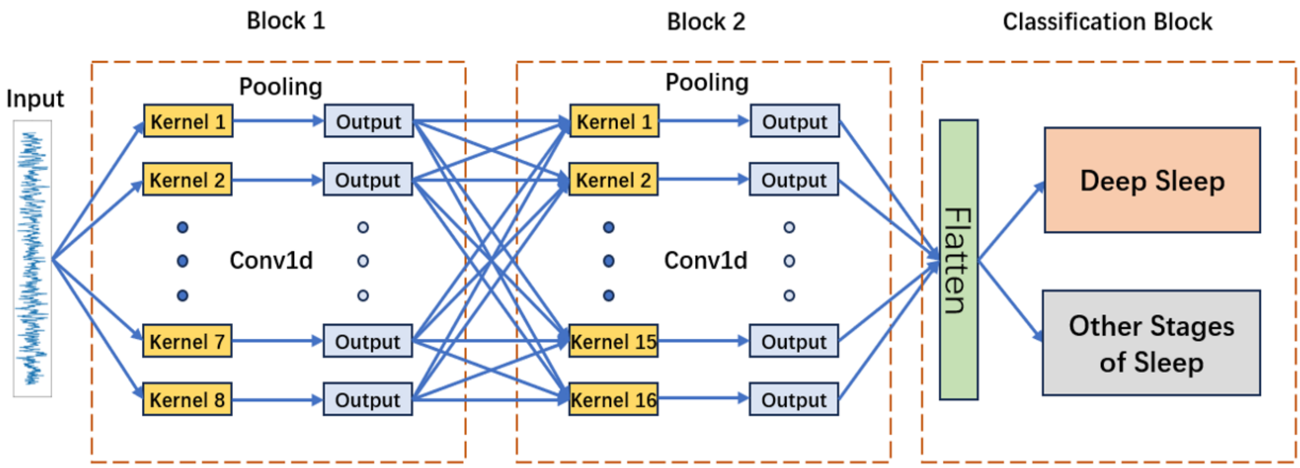

Fig. 7. Structure of EEGNet Model

K-fold Cross-validation

In machine learning model development, it is a standard approach to partition the dataset into two distinct sets: one for training and the other for testing. The test set, distinct from the training process, is not utilized during model training. Its sole purpose is to provide an unbiased evaluation of a final model fit. A frequent challenge in model training is overfitting, where the model excellently predicts training data but fails to generalize well to new, unseen data. Adjusting model parameters using the test set during the training phase would compromise the model's ability to independently evaluate unseen data, leading to skewed final evaluation results. To counteract this, part of the training data is often set aside as a validation set. This validation set is used to gauge the model's performance during the training process without compromising the test set's integrity.

The validation set, derived from the original training data, is excluded from the training phase. This exclusion is critical to objectively assess the model's predictive power on data that was not part of its learning process. The model's performance on the validation set is typically assessed using cross-validation techniques. These techniques are designed to provide a robust evaluation of the model's effectiveness on data it has not been exposed to, thereby reducing both the evaluation's bias and variance. In cross-validation, the data is segmented into K equal parts (K-Fold), with each segment alternately used as a validation set and the remaining parts as the training set. This process results in K separate models, each tested against its respective validation set (Fushiki, 2011). The performance of these models is then aggregated, usually by calculating and averaging their mean squared errors (MSE), to yield a comprehensive cross-validation error measure. This approach of 10-fold cross-validation, where the data is divided into ten parts, enhances the efficient use of data as each segment is used for both validation and training (Rodriguez et al., 2009).

In this study, the 10-fold cross-validation method, a standard variant where k equals 10, is employed. The entire dataset is evenly split into ten portions, with each portion sequentially serving as the test set and the remaining nine as the training set. This arrangement ensures each data portion is used once as a test set and nine times as part of the training set. During each of the ten cycles, the model is trained using the nine subsets and evaluated using the unique test subset. Evaluation metrics such as accuracy, precision, and mean squared error are recorded during each cycle. The culmination of this process is the calculation of the average of these metrics across all ten cycles, providing a holistic view of the model's performance. The structure of 10-fold cross-validation is shown in Fig. 8.

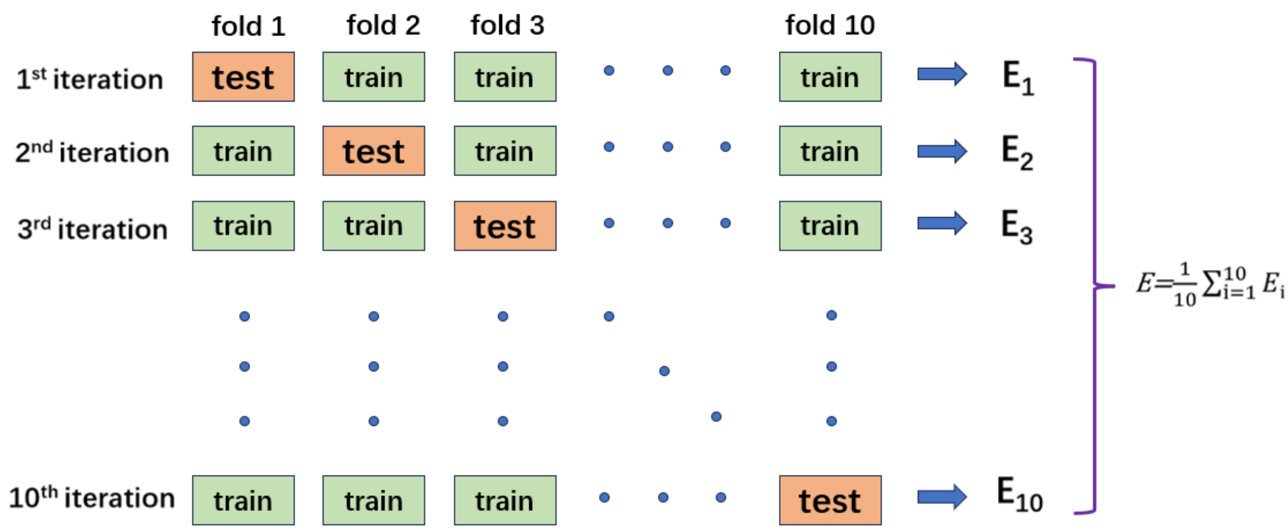

Fig. 8. K-fold cross validation (K = 10)

Data Analysis

Dataset. This article utilizes EEG data derived from two distinct sources, one being the Sleep-EDF (Extended Sleep-EDF) dataset (Kemp et al., 2000). This accessible dataset, the Sleep-EDF, offers a comprehensive collection of EEG, electrooculography (EOG), chin electromyography (EMG), along with respiration and body temperature data (our study exclusively employed EEG data) gathered from 197 individuals over an entire night. These EEG recordings were meticulously labeled by skilled technicians following the guidelines outlined in the Rechtschaffen and Kales manual ((JSSR): et al., 2001). Recorded at a 100Hz sampling rate, these annotations were conducted at 30-second intervals, yielding 3000 samples. Our research primarily focused on the SC files, acquired from a research project conducted from 1987 to 1991, which explored sleep aging effects in healthy Caucasian individuals aged between 25 and 101 years, notably in the absence of any sleep-inducing medications (Mourtazaev et al., 1995). This study involved capturing approximately two 20-hour sleep recordings of each subject at their residence over two consecutive day-night cycles (Kemp & Olivan, 2003).

Device and Experiment. In this paper, the web-based, openly accessible Sleep-EDF Dataset is employed to explore the viability of signal feature extraction for enhancing data augmentation and to analyze the performance of three specific models. The development of these models was undertaken using the Python programming language within the PyTorch framework. This task was performed on a robust workstation, which is outfitted with a 12-core Intel Core i7 3.50 GHz (5930K) processor, complemented by a GeForce RTX 2080 Ti GPU, and equipped with 128 GB of RAM for efficient processing.

The experimental data consists of a total of 12,000 entries, randomly selected from a public database. In our training approach, we employ a 10-fold cross-validation technique, where the dataset is evenly split into 10 segments. Of these, nine segments serve as the training sets, and one is utilized as the test set. To enrich our data, we apply data augmentation techniques to the EEG data within the training set, creating additional data. This augmented data is then merged with the original training set, forming an enhanced training dataset. Our training process spans 200 epochs, with each batch comprising 512 samples. During this process, we implement validation checks and retain the model weights that yield the minimum loss on the validation set. The use of 10-fold cross-validation in our methodology aims to bolster the reliability of our results. The main process of the classification is shown in Fig. 9.

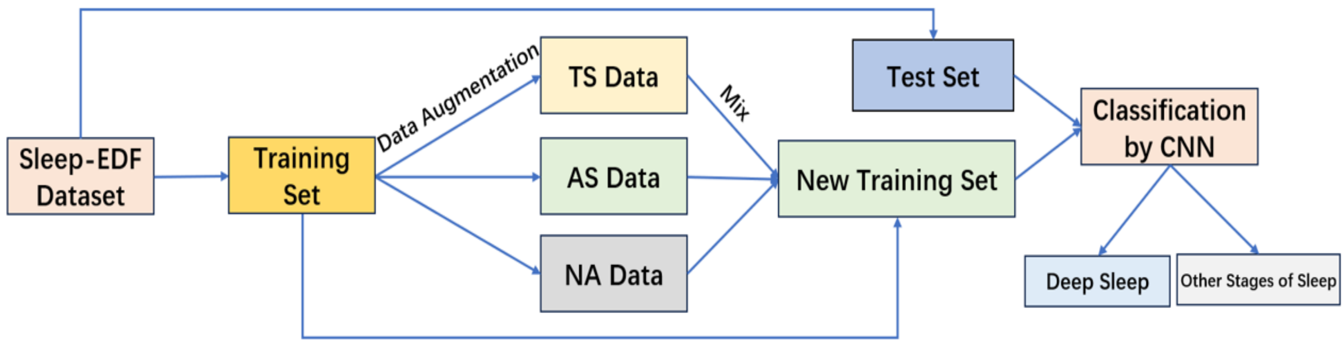

Fig. 9. Main process of classification

The model divides the entire dataset into two categories, namely deep sleep and other sleep stages, with the aim of enhancing the precision of classification and conducting a thorough analysis of the input data. Additionally, this classification system can be seamlessly adapted to other various standards, such as distinguishing between light and deep sleep phases, or categorizing sleep into REM and NREM stages.

Results

In this paper, we conducted a series of 12 experiments focused on sleep data analysis. Initially, we fed the raw sleep data independently into three neural network models: DeepConvNet, ShallowConvNet and EEGNet, for the purpose of classification learning. This allowed us to determine the accuracy of each model. Next, we employed three distinct data augmentation methods (Time Shifting, Amplitude Scaling and Noise Addition) to enhance the feature extraction process. The data produced through these augmentation techniques was then integrated with the original sleep data, creating three enriched datasets. These enhanced datasets were subsequently used to train the same models: DeepConvNet, ShallowConvNet and EEGNet, and their respective accuracies were measured. These dataset was utilized to train the three deep learning models, assessing their classification effectiveness on sleep data. The accuracies obtained using the raw sleep data and the augmented data in these deep learning models are tabulated in Table 1.

Table 1. Accuracy of raw data and augmented data with three different deep learning models

	DeepConvNet	ShallowConvNet	EEGNet
Raw Data	94.12%	92.15%	93.38%
Raw Data + TS Data	95.05%	92.26%	94.46%
Raw Data + AS Data	94.75%	92.87%	94.32%
Raw Data + NA Data	94.60%	92.31%	93.50%

While the outcomes didn't achieve the optimal accuracy in pinpointing deep sleep stages, our experiments highlight that incorporating a variety of data augmentation techniques across several deep learning models enhances their accuracy beyond the use of unprocessed data. This method of data augmentation in the preprocessing phase is notably more efficient and expedient than alternatives like EMD or entropy techniques.

Our experiments demonstrate that employing three different data augmentation methods for feature extraction significantly enhances the performance of all three evaluated deep learning models. The combined application of these methods resulted in a notable boost in accuracy over models that relied solely on raw data. Additionally, there is considerable potential for refining the accuracy further by fine-tuning the structural parameters of the neural networks and optimizing the specifics of the data augmentation process. This approach opens up new avenues for advancing the precision of sleep stage classification, paving the way for more effective and nuanced sleep analysis.

Conclusions

In recent years, the rapid development of machine learning has garnered widespread attention, with increasing efficiency and accuracy, opening new possibilities in the field of signal recognition. Locating deep sleep stages essentially involves recognizing different signals, and relying on machine learning for this task is a promising research direction. In this paper, we devised a new method using a data augmentation feature extraction algorithm to provide more information to CNNs, thereby achieving more accurate models. We also presented and compared the performance of three different models (DeepConvNet, ShallowConvNet and EEGNet) in EEG data classification. Based on our experimental results, we found that three data augmentation methods (Time Shifting, Amplitude Scaling and Noise Addition) significantly improved model accuracy for various deep learning models. Compared to the latest classification results, these networks still have room for improvement in accuracy. Further tuning of parameters and optimizing the CNN structure should lead to higher accuracy levels. Additionally, in subsequent experiments, it is worthwhile to try combining more transformations such as mirroring, elastic deformation and frequency rotation, or using more channels of other feature extraction methods to enhance accuracy.

Previous research has already demonstrated the effectiveness of deep learning methods in accurately and stably identifying deep sleep stages in clinical practice. Moreover, the training and classification process of the entire system is more efficient and cost-effective than manual identification. Once trained, the model can be used long-term, offering practical application value in real life. Furthermore, it is reasonable to attempt to train classification models in other areas of health assessment based on biosignal frequency domain features, thereby achieving improved recognition and cognitive accuracy.

Sleep is closely related to everyone's life, yet deep sleep is crucial for bodily repair and recovery, immune system strengthening, memory consolidation and enhancement of learning abilities. Therefore, we hope to monitor the deep sleep stages in our sleep using the aforementioned methods, applying scientific approaches to improve the proportion of deep sleep in everyone's sleep cycle.

Acknowledgments

This work was supported by JSPS KAKENHI 20H04249.

Conflicts of Interest

The author has no conflict of interest about anything in this article.

References

Anicet Zanini, R., & Luna Colombini, E. (2020). Parkinson's disease emg data augmentation and simulation with dcgans and style transfer. *Sensors*, *20*(9), 2605. https://doi.org/10.3390/s20092605

Barnes, C. M., Lucianetti, L., Bhave, D. P., & Christian, M. S. (2015). "you wouldn't like me when i'm sleepy": Leaders' sleep, daily abusive supervision, and work unit engagement. *Academy of Management Journal*, *58*(5), 1419–1437. https://doi.org/10.5465/amj.2013.1063

Basics, B. (2021). Brain basics: Understanding sleep—national institute of neurological disorders and stroke.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

Caporale, A., Lee, H., Lei, H., Rao, H., Langham, M. C., Detre, J. A., Wu, P.-H., & Wehrli, F. W. (2021). Cerebral metabolic rate of oxygen during transition from wakefulness to sleep measured with high temporal resolution oxflow mri with concurrent eeg. *Journal of Cerebral Blood Flow & Metabolism*, *41*(4), 780–792. https://doi.org/10.1177/0271678X20919287

Carskadon, M. A., Dement, W. C., et al. (2005). Normal human sleep: An overview. *Principles and practice of sleep medicine*, *4*(1), 13–23.

Chen, Z., Cao, J., Cao, Y., Zhang, Y., Gu, F., Zhu, G., Hong, Z., Wang, B., & Cichocki, A. (2008). An empirical eeg analysis in brain death diagnosis for adults. *Cognitive Neurodynamics*, *2*, 257–271. https://doi.org/10.1007/s11571-008-9047-z

Cirelli, C., & Tononi, G. (2008). Is sleep essential? *PLoS biology*, *6*(8), e216.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273–297. https://doi.org/10.1007/BF00994018

De Gennaro, L., & Ferrara, M. (2003). Sleep spindles: An overview. *Sleep medicine reviews*, *7*(5), 423–440.

Eren, L., Ince, T., & Kiranyaz, S. (2019). A generic intelligent bearing fault diagnosis system using compact adaptive 1d cnn classifier. *Journal of Signal Processing Systems*, *91*, 179–189. https://doi.org/10.1007/s11265-018-1378-3

Fushiki, T. (2011). Estimation of prediction error by using k-fold cross-validation. *Statistics and Computing*, *21*, 137–146.

Guresen, E., & Kayakutlu, G. (2011). Definition of artificial neural networks with comparison to other networks. *Procedia Computer Science*, *3*, 426–433. https://doi.org/10.1016/j.procs.2010.12.071

Ho, T. K., et al. (1995). Proceedings of 3rd international conference on document analysis and recognition. *IEEE*, 278–282.

Hobson, J. A., & Pace-Schott, E. F. (2002). The cognitive neuroscience of sleep: Neuronal systems, consciousness and learning. *Nature Reviews Neuroscience*, *3*(9), 679–693.

Iber, C. (2007). The aasm manual for the scoring of sleep and associated events: Rules, terminology, and technical specification. *(No Title)*.

Iwana, B. K., & Uchida, S. (2021). An empirical survey of data augmentation for time series classification with neural networks. *Plos one*, *16*(7), e0254841.

(JSSR): S. C. C. O. T. J. S. O. S. R. S., Hori, T., Sugita, Y., Koga, E., Shirakawa, S., Inoue, K., Uchida, S., Kuwahara, H., Kousaka, M., Kobayashi, T., et al. (2001). 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. *Psychiatry and clinical neurosciences*, *55*(3), 305–310. https://doi.org/10.1046/j.1440-1819.2001.00810.x

Kemp, B., & Olivan, J. (2003). European data format 'plus'(edf+), an edf alike standard format for the exchange of physiological data. *Clinical neurophysiology*, *114*(9), 1755–1761. https://doi.org/10.1016/S1388-2457(03)00123-8

Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A., & Oberye, J. J. (2000). Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the eeg. *IEEE Transactions on Biomedical Engineering*, *47*(9), 1185–1194. https://doi.org/10.1109/10.867928

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, *15*(5), 056013. https://doi.org/10.1088/1741-2552/aace8c

Li, B., & Cao, J. (2023). Classification of coma/brain-death eeg dataset based on one-dimensional convolutional neural network. *Cognitive Neurodynamics*, 1–12. https://doi.org/10.1007/s11571-023-09942-2

Mekruksavanich, S., & Jitpattanakul, A. (2022). Cnn-based deep learning network for human activity recognition during physical exercise from accelerometer and photoplethysmographic sensors. In *Computer networks, big data and iot: Proceedings of iccbi 2021* (pp. 531–542). Springer. https://doi.org/10.1007/978-981-19-0898-9_42

Michel, C. M., & Murray, M. M. (2012). Towards the utilization of eeg as a brain imaging tool. *Neuroimage*, *61*(2), 371–385. https://doi.org/10.1016/j.neuroimage.2011.12.039

Mourtazaev, M., Kemp, B., Zwinderman, A., & Kamphuisen, H. (1995). Age and gender affect different characteristics of slow waves in the sleep eeg. *Sleep*, *18*(7), 557–564. https://doi.org/10.1093/sleep/18.7.557

Niedermeyer, E., & da Silva, F. L. (2005). *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Ohayon, M. M. (2002). Epidemiology of insomnia: What we know and what we still need to learn. *Sleep medicine reviews*, *6*(2), 97–111.

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2009). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE transactions on pattern analysis and machine intelligence*, *32*(3), 569–575. https://doi.org/10.1109/TPAMI.2009.187

Roots, K., Muhammad, Y., & Muhammad, N. (2020). Fusion convolutional neural network for cross-subject eeg motor imagery classification. *Computers*, *9*(3), 72. https://doi.org/10.3390/computers9030072

Sanei, S., & Chambers, J. A. (2013). *Eeg signal processing*. John Wiley & Sons.

Sarangi, S., Sahidullah, M., & Saha, G. (2020). Optimization of data-driven filterbank for automatic speaker verification. *Digital Signal Processing*, *104*, 102795. https://doi.org/10.1016/j.dsp.2020.102795

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, *6*(1), 1–48. https://doi.org/10.1186/s40537-019-0197-0

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278.

Xia, M., Zhao, X., Deng, R., Lu, Z., & Cao, J. (2023). Eegnet classification of sleep eeg for individual specialization based on data augmentation. *Cognitive Neurodynamics*, (In press).

Zhao, X., Zhao, Q., Tanaka, T., Solé-Casals, J., Zhou, G., Mitsuhashi, T., Sugano, H., Yoshida, N., & Cao, J. (2023). Classification of the epileptic seizure onset zone based on partial annotation. *Cognitive Neurodynamics*, *17*(3), 703–713. https://doi.org/10.1007/s11571-022-09857-4