A Two-Stage Learning Convolutional Neural Network for Sleep Stage Classification Using a Filterbank and Single Feature

MEHDI ABDOLLAHPOUR, TOHID YOUSEFI REZAI, ALI FARZAMNIA, ISMAIL SAAD

ABSTRACT
Sleep is an essential process for the body that helps to maintain its health and vitality. The first stage in the diagnosis and treatment of sleep disorders is sleep staging. Due to the complications in manual sleep staging by the physician, computer-aided sleep stage classification algorithms are gaining attention. In this study, a novel approach was introduced to extract distinctive representations from single-channel EEG signal for automatic sleep staging. Standard deviation as a single feature was extracted from the frequency subbands of EEG, which gave a comprehensive understanding of the signal and its activity within various frequency ranges for different sleep stages. The features formed the input space of the proposed two-stream convolutional neural network (CNN) for classification and two-stage learning was used to train the model that achieved improvements in terms of accuracy, reliability and robustness against traditional classifiers and conventional training method of the neural networks. For the performance evaluation, three well-known benchmark datasets including Sleep EDF, Sleep EDFx and DREAMS Subject were used. The proposed algorithm by utilizing simple and effective methods improved sleep stage classification results by achieving an overall accuracy of 93.48%, 93.14% and 83.55%, respectively. The introduced framework in this study has great potential for practical implementation on a home-based sleep staging device.

INDEX TERMS Sleep stage classification, filterbank, single channel EEG, single feature, convolutional neural network, 2-stage learning.

I. INTRODUCTION
Proper overnight sleep is crucial for the body to recover from the tiredness of the day. Any abnormality in the healthy sleep cycle can cause short-term and long-term problems for the body. On the other hand, sleep problems can be a sign of a pre-existing condition. As suggested in epidemiological studies, insomnia is a common symptom of depression, and on the other hand, other sleep problems such as difficulties in falling asleep, frequent awakenings and obstructive sleep apnea can directly affect a person’s everyday life [1].

The first stage in the diagnosis of sleep related diseases is sleep stage classification. According to the American Academy of Sleep Medicine (AASM) rules [2] there are five different sleep stages including Wake, N1, N2, N3, and Rapid Eye Movement (Rem). AASM rules include a set of guidelines for sleep experts to help them classify sleep stages using Polysomnography (PSG) signals. PSG records multiple physiological signals such as Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG), and also heart rate, breathing and oxygen level during overnight sleep. The sleep expert or physician by inspecting these signals assigns a sleep stage to every 30-second epoch of the signals. The output of this process will be a cycle of sleep stages called hypnogram. The hypnogram can be assessed by the physician to diagnose sleep disorders. [3].

Manual sleep staging takes place in a sleep clinic and the
patient should spend one or two nights in the clinic. Some people may feel uncomfortable when sleeping in new places for the first time and may have difficulty falling asleep [4]. On the other hand, manually checking the PSG signals by the physician to detect changes in the sleep stage is not an efficient way to classify sleep stages. Manual sleep staging is a time-consuming and costly task, and the results can be affected by human error. As a result, many studies were conducted to overcome the mentioned problems. These studies introduce algorithms for automatic sleep stage classification, which can acquire hypnogram from the PSG signals without a need for a sleep expert. Generally, automatic sleep staging methods use single-channel EEG or a combination of EEG, EOG, and EMG signals for classification [5–7].

Automatic sleep stage classification algorithms include an extensive area of machine learning methods [8]. These algorithms can be divided into two groups including, classical machine learning methods and deep learning-based methods. The first stage in classical machine learning methods is feature extraction. Various signal processing and feature extraction methods have been developed to extract unique characteristics of the signals for classification. Common features that have been extracted from PSG signal for sleep staging include time domain, frequency domain, and non-linear features [9–12]. Feature selection is essential to achieve acceptable recognition for different sleep stages. In the next step, extracted features provide the input space for a classifier. Well-known classifiers such as k-nearest neighbor (KNN) [13], decision tree (DT) [14], random forest (RF) [11], and support vector machine (SVM) [15] are commonly used in classical sleep staging algorithms.

By introducing novel deep learning frameworks and the development of neural network structures, deep learning methods gained more attention in sleep staging studies. Neural networks can perform significantly flexible for different data types and outperform traditional classifiers in different scenarios. Different types of neural networks have been evolved for sleep stage classification. One of the most popular deep learning models is recurrent neural networks (RNN) which uses current input and previous input of the network to determine the current state of the network. RNN can handle sequential datasets like sleep data whose current sleep stage is related to the previous stage. A specific type of RNNs called long short-term memory (LSTM) [16] is popular in sleep studies [5], [17]. Convolutional neural networks (CNN) by achieving promising results for classification of different datasets and a variety of data types such as signal, image, and video, is one of the commonly used classifiers [18]–[20]. Consecutive non-linear functions and adaptive convolutional filters in different layers of a CNN make it possible to design a flexible model to extract complex features of the input data and make good predictions.

Zhu et al. [21] proposed an attention-based convolutional neural network that helped convolutional layers in feature extraction. Attention layers are used to resemble an RNN for detecting sequential information of sleep stages. Chambon et al. [22] introduced a deep learning architecture for temporal classification of sleep stages using multi-channel EEG, EOG, and EMG signals. The proposed model in this study is based on performing spatial filtering on different channels before convolutional layers and separate pipelines for EEG/EOG and EMG signals. In another study conducted by Cui et al. [23] multi-channel EEG time series are used for sleep staging without extracting hand-crafted features. Raw EEG signals are segmented by a method called fine-grained segments and fed to a multi-stream CNN. Li et al. [24] proposed an end-to-end CNN for sleep staging using raw EEG signals. The main idea of this work is to classify three subsequent EEG epochs using a novel CNN architecture based on squeeze-and-excitation block and micro-network.

Most of the deep learning-based studies concentrate on using only raw EEG for classification, therefore valuable hand-crafted features that were shown to be beneficial for sleep staging are neglected in these methods. On the other hand, raw EEG is a complex, non-stationary, and non-linear time series that takes too much effort to design a deep learning structure that can capture its information. Deep learning frameworks have the potential to perform perfectly when there is enough data to train the network. However, in the case of sleep staging because of the difficulties of data recording and labeling the data, there is a limited number of data samples. Therefore, this point needs special attention to prevent inconsistent results with low generalizability [25].

Considering the mentioned problems, we introduced a sleep staging framework to process the single-channel EEG signal and extract hand-crafted features. A single feature is extracted from the subbands of the EEG signal, which is decomposed by a filterbank. Besides, a two-stage learning CNN is introduced to classify extracted features. We are aiming to demonstrate that in the case of the sleep stage classification it is possible to propose a simple solution to distinguish the main characteristics of the EEG for different sleep stages. Furthermore, training the CNN in two stages can help to achieve consistent and reproducible results with high accuracy even with databases with small size like sleep staging databases. The outline of the proposed method is shown in Fig. 1. The key contributions of the current work are as follows.

1) A novel framework is introduced for extracting hand-crafted features from EEG signal using a filterbank and a single statistical feature. A simple and straightforward approach to extract distinctive representations of EEG signal. This framework is designed to obtain only useful information from EEG with minimum redundancy and complications.

2) A new method is proposed for training the CNN in two stages that improves the performance of the CNN in terms of, consistency, reproducibility, reliability, and accuracy without adding extra computational complexity to the problem.

3) The proposed algorithm is evaluated using three benchmark datasets for sleep staging including Sleep EDF,
Sleep EDFx and DREAMS Subject. Promising results are achieved for different datasets that show the high generalizability of the model. Furthermore, the results show improvements in sleep stage classification in comparison to the state-of-the-art studies.

The remainder of the study is organized as follows. In Section II, benchmark datasets that are used for model evaluation are introduced and the proposed method is described in detail. In Section III, various experiments are designed to evaluate the performance of the algorithm in different situations. In Section IV proposed algorithm is discussed from different points of view. Section V draws a conclusion on the current work.

II. MATERIAL AND METHODS
A. DATASETS
1) Sleep-EDF
Sleep-EDF is a publicly available dataset that is accessible in the PhysioNet database [26], [27]. Sleep-EDF is the most commonly used dataset for the evaluation of sleep staging algorithms. This dataset is composed of two different data recording scenarios including sleep telemetry (ST) and sleep cassette (SC). In SC recording, physiological signals of the 4 subjects are recorded during their daily lives for 24 hours using a portable recording device. ST recordings are recorded in the hospital during the night and include the physiological signals of 4 healthy subjects who had mild difficulty falling asleep. The subjects were healthy males and females aged between 21 and 36. PSG signals including EEG, EOG and EMG are recorded continuously from subjects during sleep and sleep specialist assigned a sleep stage to each 30-second epoch of signals according to the R & K rule [28]. The acquired hypnogram contains Wake, REM, S1, S2, S3, S4 sleep stages and also movement time and unscored stages, which should be trimmed. In this study, an EEG signal from Fpz-Cz channel with a sampling frequency of 100 Hz is used for classification. Furthermore, S4 and S3 sleep stages are merged into N3 according to the AASM rule and the epochs that are related to the movement time and unscored are removed from hypnogram and EEG signal.

2) Sleep EDFx
Sleep EDF expanded (EDFx) is the expanded version of the Sleep-EDF dataset. More subjects participated in the recording of the Sleep-EDFx. The same scenario as in the Sleep-EDF is performed on the current dataset to acquire ST and SC recordings. However, in Sleep EDFx, ST recordings were obtained to examine temazepam effects on sleep from 22 healthy male and female Caucasians and SC recordings were obtained to examine age effects on sleep from healthy Caucasians aged between 25 and 101.

3) DREAMS Subject Database
DREAMS Subject database is another publicly available benchmark dataset for the assessment of sleep staging algorithms. It consists of overnight polysomnographic signals of 20 healthy subjects (4 male and 16 female) aged between 20 and 65. The dataset was collected at the Sleep Laboratory of the André Vésale hospital in Belgium. Two EOG channels, three EEG channels (CZ-A1 or C3-A1, FP1-A1 and O1-A1) and one submental EMG channel were recorded from each subject at a sampling frequency of 200 Hz. Each 30-second epoch of signals was labeled according to the AASM rule as Wake, N1, N2, N3, REM, and unknown sleep stage. Detailed information about the datasets is mentioned in Table 1.

B. FEATURE EXTRACTION
In the manual sleep staging methods, the frequency activity of the EEG signal is the most powerful indicator for sleep specialists to recognize sleep stages. According to the AASM rule, two important factors are described for sleep experts to help them distinguish different sleep stages, including frequency content and the amplitude of the EEG signal. An EEG signal can be decomposed into different waves so that each of these waves is active in a specific frequency range. The frequency range of different waves and the presence of these waves in different sleep stages [29]–[32]. Various studies have tried to reveal the relationship between the frequency content of the EEG signals and sleep stages. These studies have been trying to utilize frequency domain, time domain and non-linear features to detect different sleep stages. However, in these works due to a large number of involved parameters and complexity of the algorithm, it is not possible to discuss the effectiveness of frequency-domain features in recognition of different sleep stages. On
### TABLE 1. Description of the evaluation datasets.

| Dataset     | No. of subjects | EEG channel   | Epochs (Total) | Wake (%) | N1(%) | N2(%) | N3(%) | REM(%) |
|-------------|-----------------|---------------|----------------|----------|-------|-------|-------|--------|
| Sleep-EDF   | 8               | Fpz-Cz        | 15188          | 53.0     | 4.0   | 23.8  | 8.6   | 10.6   |
| Sleep-EDFx  | 20              | Fpz-Cz        | 36963          | 54.7     | 3.4   | 23.0  | 9.3   | 9.6    |
| DREAMS      | 20              | Cz-A1         | 20242          | 17.6     | 7.3   | 40.8  | 19.4  | 14.9   |

### TABLE 2. Characteristics of EEG epochs for each sleep stage including the presence of various waves in the EEG epochs and dominant frequency range. LAMF stands for low amplitude mixed frequency.

| Sleep Stage | Present waves and characteristics          | Frequency range (Hz) |
|-------------|--------------------------------------------|----------------------|
| Wake        | Alpha with eyes open                        | (8 - 13)             |
|             | LAMF with eyes closed                       | (20 - 50)            |
|             | Beta, Gamma                                | (0.5 - 2)            |
| N1          | Theta, LAMF, Vertex sharp                   | (4 - 7)              |
| N2          | Theta, K-complexes                         | (4 - 7)              |
|             | Sleep spindles with high amplitude         | (11 - 16)            |
| N3          | Slow waves with high amplitude              | (0.5 - 3)            |
|             | Sleep spindles                             |                      |
| REM         | Rapid eye movement                         | (0.1 - 1)            |
|             | Theta                                      | (4 - 7)              |
|             | Slow alpha                                 | (6 - 10)             |
|             | Saw tooth waves                            | (2 - 6)              |

On the other hand, most of the frequency domain features are based on frequency bands such as delta, theta, and alpha that are not good at providing frequency information with acceptable resolution. From Table 2, we can see the same frequency bands are present in different sleep stages. Therefore, band powers are not the best features for distinguishing different sleep stages. Due to the mentioned problems, we propose a framework to extract features from single-channel EEG signal that provide distinctive frequency and amplitude information for different sleep stages.

1) Preprocessing

To generalize the method to different datasets, a baseline preprocessing method is considered. First, all of the signals are downsampled to 100 Hz using the polyphase filter because only the 0 to 50 Hz frequency information is needed for further processing. Then, the recorded overnight EEG of each subject is centered and normalized by standard deviation. It helps to achieve a good representation of the signals in the feature extraction steps, and it is also necessary to have normalized input for a convolutional neural network (CNN) to have a faster training process.

According to the AASM rule, each 30-second epoch of EEG signal should be associated with a specific sleep stage. Furthermore, in the AASM rule in many cases, to recognize each epoch of the EEG signal, the first half and second half of the epoch can be inspected for the presence of specific wave and frequency activity. Therefore, in the proposed method, 30-second epochs are divided into two 15-second subepochs for further processing. As a result, 15-second subepochs of EEG with 1500 data points are prepared for filtering and feature extraction.

2) Filterbank

The first step is the decomposition of the input signal into frequency subbands. The proposed filterbank consists of 50 bandpass filters with a frequency range of 1 Hz without overlap that spans the 0 to 50 Hz range. The filters are infinite impulse response Butterworth with an order of 3. Choosing a higher order for the bandpass filter will produce sharper frequency boundaries. However, it may cause unstable results and fluctuating signals at the output. In the case of sleep staging, selecting a lower order for the filterbank and using slick bandpass frequency boundaries showed meaningful improvements in the results. Afterward, each subepoch is passed through the filterbank separately and 50 subbands of EEG signal are produced as output. In the proposed setup for filtering, an EEG signal with higher frequency activity in a specific frequency range will have active subbands for the corresponding frequencies after filtering while other subbands are suppressed.

3) Standard Deviation

In the feature extraction step, a common approach in signal classification studies is to use a combination of different features from various domains such as time domain, frequency domain, and non-linear features. These methods obtained acceptable results in sleep staging. However, these features are not specifically designed to distinguish different sleep stages. In the proposed method, we used standard deviation as a single feature to extract characteristics of EEG subbands. Standard deviation is a measure of the dispersion of data around its mean, which can be calculated as,
where $\mu$ is the input signal, $N$ is the number of data points and $\overline{\mu}$ is the mean of the input signal.

Standard deviation will be calculated for each subband of the first and second subepoch. As a result, we will have 49 features corresponding to 1Hz to 49Hz frequencies for each 15-second subepoch for the classification. Due to power line interference, the standard deviation of the 50Hz frequency subband was removed from the features. Fig. 2 shows the extracted feature from one epoch of the EEG signal.

\[
\text{std}(x) = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{\mu})^2}{N}}.
\]  

(1)

2) Two-Stage Learning

Convolutional neural networks are based on learning representations of input data using convolutional layers and making predictions using fully connected layers. Each of these sets of layers includes a large number of trainable parameters, therefore a large number of training data is needed to achieve satisfactory results [33]. In the case of sleep stage classification, datasets with large sizes are not available and this leads to inconsistent results in classifying sleep stages with CNN. Additionally, convolutional layers and fully connected layers pursue a different goal in the learning process which is feature extraction and classification, respectively. However, each parameter of the different layers gets an update at the same time in each epoch regardless of the type of the layers which can lead to a slow pace of learning and divergence in the learning process.

To overcome the mentioned problems, in our last work [34] we proposed a method for training a two-stream CNN such that each stream of convolutional layers was trained and frozen in the separate sub-networks, and transferred to the main model.

Each layer of a feedforward CNN consists of a number of neurons with learning capability. Each neuron has weights and a bias that hold the learned information during the training process. For the training, an optimizer computes the gradients of the loss function. Then, through the backpropagation method, gradients will be used to update weights and biases for all of the neurons from the last layer toward the first layer. The described process is the conventional way of training CNNs. In this process, all of the parameters of the model will be updated in each iteration of the training.

The proposed framework for training takes place in two stages. The hyperparameters of the model are manually tuned at first and are not changed during the training. The model includes 6 convolutional layers and 3 fully connected (FC) layers with a total of 390461 trainable parameters. Each neuron in a layer of a CNN is connected to all of the neurons in the previous layer by a trainable parameter. This parameter can be updated in the training to pass useful information to the next layer.

The Adam optimizer [35] with a learning rate of 0.0003 is used for the training. At the first stage of the two-stage learning, the CNN is trained in the conventional way, and all of the model parameters are updated for 100 epochs. At this stage, convolutional layers are adapting themselves to the input data and learning the representations of it to extract distinctive information. The outputs of convolutional layers from two streams are concatenated as activations for the FC layers. Meanwhile, FC layers are learning to recognize these
activations as the respective sleep stages. At the end of the first stage’s training, the model is not overfitted and shows an acceptable performance in classifying sleep stages. At the second stage, the parameters of the convolutional layers are frozen, so their values will not change during training. At this stage, train data is fed to the model for another 100 epochs for FC layers to continue learning. In the second stage, the data representations do not change during the training, and this gives the model a chance to train FC layers at their full capacity for the classification.

In this approach, the training of convolutional layers will be stopped before the end of the training. After 100 epochs of training, only the weights of FC dense layers will need to be updated, resulting in a faster training process for the model.

III. EXPERIMENTS
In this section, the performance of the proposed method will be investigated from different points of view. Different experiments are designed to make sure the algorithm will perform in different situations on different datasets. In the following, the important criteria for performance assessment of sleep staging algorithms are introduced.

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}
\]
\[
\text{Specificity} = \frac{TN}{TN + FP} \tag{3}
\]
\[
\text{Precision} = \frac{TP}{TP + FP} \tag{4}
\]
\[
F1 - \text{Score} = \frac{2TP}{2TP + FP + FN} \tag{5}
\]

Where TP, TN, FP, FN stand for True positive, True negative, false positive, and false negative, respectively.

In addition to the accuracy as a measure for the performance of the algorithm, Cohen’s kappa [36] is also used to calculate the agreement between expert scored sleep stages and algorithm scored sleep stages. Generally, Cohen’s kappa is known to be a robust measure in comparison to the accuracy because it takes into account the possibility of the agreement between raters occurring by chance.

A. EXPERIMENT 1: THE LEARNING CURVE OF THE PROPOSED CNN
This experiment is designed to evaluate the performance of the proposed CNN architecture and two-stage learning process based on the accuracy and loss of the model during the training. Two scenarios are considered. In the first one, the CNN is trained for 200 epochs in the traditional way and in the other scenario the proposed method for 2-stage training is implemented and test loss and test accuracy in both scenarios are considered for comparison. Train and test data samples are selected the same in both scenarios where the test data and train data include 20% and 80% of the whole dataset, respectively.

The results are shown in Fig. 4. Each row represents a specific dataset including DREAMS, Sleep EDF, and Sleep EDFx. As it can be seen, in 2-stage learning a significant improvement occurred after epoch 100. In comparison to the 1-stage learning, the fluctuations in accuracy curves decreased and the loss value has finely diverged to its optimum value. For a quantitative comparison, mean, standard deviation, minimum and maximum of the accuracy and loss curves are mentioned in Table 3 for epoch 100 to 200. In 2-stage learning, the results are more consistent, reliable, and reproducible for different runs. In addition, the results show an improvement in accuracy values.

B. EXPERIMENT 2: PERFORMANCE COMPARISON OF THE PROPOSED METHOD AND TRADITIONAL CLASSIFIERS
This experiment is designed to illustrate the effectiveness of the proposed CNN in comparison to the other classical approaches in feature classification. Four cases includ-
TABLE 3. Compared statistics of test accuracy for 1-stage and 2-stage learning. The values are calculated for the last 100 epochs of training.

| Dataset   | Mean ±std | Min  | Max  |
|-----------|-----------|------|------|
| 2-stage learning |
| DREAMS    | 83.3±0.18 | 82.9 | 83.7 |
| EDF       | 93.4±0.20 | 92.5 | 93.8 |
| EDFx      | 93.0±0.11 | 92.7 | 93.3 |
| 1-stage learning |
| DREAMS    | 82.6±0.6  | 79.7 | 83.7 |
| EDF       | 92.9±0.45 | 91.4 | 93.6 |
| EDFx      | 92.4±0.43 | 91.1 | 93.2 |

In this experiment, a highly imbalanced dataset with more epochs in the wake stage and few epochs in the N1 stage. Therefore, we can say classifiers with a high recognition rate for the wake stage in comparison to the other stages, are biased towards the wake stage. On the other hand, the low number of N1 training epochs caused LR, SVM, and GNB to be completely unable to recognize the N1 sleep stage.

C. EXPERIMENT 3: PERFORMANCE COMPARISON AGAINST BASELINE WORKS

In this section, a comprehensive evaluation is conducted on the proposed method based on different criteria using various datasets. The performance of the algorithm will be compared to the state-of-the-art and baseline methods for automatic sleep stage classification. Each of the datasets is randomly split into train and test sets in which test and train samples comprise 20% and 80% of the total dataset, respectively. For a small dataset such as Sleep EDF, selecting data samples for the test and train set can affect the final result. Therefore, to have a fair comparison between the performance of the proposed method and other methods, the reported result for the Sleep EDF dataset is the average of five different runs.

Table 5 is the confusion matrix of the algorithm on the Sleep EDF dataset. Each row in the confusion matrix represents the epochs classified by the sleep expert for a specific stage, where each column includes epochs classified by the algorithm. The diagonal elements of the confusion matrix are where the expert’s and algorithm’s opinions meet, which represent the sensitivity for a specific sleep stage. Per-stage specificity, F1 score, and precision are also mentioned in the bottom lines of the table. These measures help to have a comprehensive and robust evaluation of the proposed algorithm.

As can be seen from Table 5, the algorithm is succeeded...
to recognize Wake, N2, N3, and REM with a sensitivity of 92.3%, 93.3%, 88% and 88.5%, respectively. On the other hand, due to the imbalanced classes of sleep datasets, the obtained sensitivity for class N1 is 55.1% which is lower with respect to the other classes. However, F1-score and precision can be considered better criteria in comparison to the sensitivity as it takes into account the unbalancedness of the dataset.

Table 6 and Table 7 illustrate the confusion matrix and a detailed description of the obtained results from classifying Sleep EDF and DREAMS Subjects datasets. The sensitivity values for the DREAMS Subjects dataset are lower in comparison to the Sleep EDF and Sleep EDFx datasets. The reasons for this are the differences in data acquisition system, EEG channel, and distribution of data samples over different sleep stages. However, the overall performance of the method is similar in classifying different sleep stages for three benchmark datasets that shows the consistency of the proposed method in the face of different datasets.

### Table 4.
Performance comparison between the proposed method and traditional classifiers for 2-stage to 5-stage classification in terms of per-class sensitivity. The highest sensitivity values per class are shown in bold.

|       | 2-class | 3-class | 4-class | 5-class |
|-------|---------|---------|---------|---------|
| W     | S       | W       | NREM    | R       |
| LDA   | 91.3    | 95.1    | 94.5    | 84.9    | 85.0    | 94.9    | 28.6    | 89.9    | 84.9    |
| RF    | 97.6    | 97.0    | 98.5    | 94.4    | 94.4    | 80.4    | 98.4    | 35.3    | 94.6    |
| KNN   | 94.7    | 92.4    | 95.4    | 86.2    | 71.0    | 95.5    | 34.5    | 89.3    | 74.4    |
| LR    | 91.3    | 95.3    | 93.9    | 88.1    | 69.0    | 95.0    | 9.2     | 92.6    | 80.4    |
| DT    | 97.6    | 97.1    | 98.2    | 94.3    | 78.9    | 98.3    | 37.8    | 96.6    | 84.3    |
| SVM   | 93.4    | 96.3    | 94.4    | 88.1    | 81.3    | 95.2    | 1.6     | 93.5    | 85.0    |
| GNB   | 58.2    | 96.3    | 57.8    | 54.7    | 87.6    | 57.5    | 15.1    | 71.0    | 86.7    |
| Proposed | 98.1    | 98.2    | 98.2    | 95.0    | 89.8    | 98.3    | 49.6    | 97.5    | 86.8    |

### Table 5.
Confusion matrix and per-class performance of the proposed algorithm for the 5-stage classification on Sleep EDF dataset.

|       | Wake | N1   | N2   | N3   | REM  |
|-------|------|------|------|------|------|
| Expert scored | 1592 | 17   | 2    | 1    | 6    |
| Wake   | 98.3%| 1.0% | 0.1% | 0.3% | 0.4% |
| N1     | 13   | 67   | 16   | 0    | 24   |
| 11.3%  | 55.1%| 13.3%| 0.3% | 20.0%|
| N2     | 3    | 5    | 673  | 25   | 15   |
| 0.5%   | 0.7% | 93.3%| 3.5% | 2.1% |
| N3     | 2    | 1    | 26   | 220  | 0    |
| 0.9%   | 0.5% | 10.5%| 88.0%| 0.0% |
| REM    | 6    | 9    | 21   | 0    | 286  |
| 1.8%   | 2.9% | 6.7% | 0.0% | 88.5%|

Specificity (%) | 98.2 | 98.9 | 97.1 | 99.0 | 98.3
F1 Score (%)     | 98.4 | 60.5 | 92.1 | 88.6 | 87.3
Precision (%)    | 98.4 | 67.2 | 91.0 | 89.2 | 86.1

Specificity (%) | 98.2 | 99.2 | 96.1 | 99.0 | 98.4
F1 Score (%)     | 98.7 | 45.8 | 90.3 | 88.0 | 86.2
Precision (%)    | 98.5 | 60.8 | 88.0 | 90.0 | 85.4

The final step in the assessment of the algorithm is to find the place where the proposed algorithm stands between the sleep staging algorithm. Baseline and state-of-the-art studies in the single-channel sleep stage classification area are compared based on accuracy and Cohen’s kappa in Table 8. The comparison is based on the subject non-independent paradigm and the data samples of each subject can be used for training and testing of the model. It can be seen, the proposed algorithm outperformed other works in classifying...
TABLE 7. Confusion matrix and per-class performance of the proposed algorithm for the 5-stage classification on Dreams Subject dataset.

| Stages | Wake | N1  | N2  | N3  | REM  |
|--------|------|-----|-----|-----|------|
|        |      |     |     |     |      |
| Expert scored |      |     |     |     |      |
|         | 661  | 59  | 22  | 6   | 6    |
| Specificity(%) | 92.8% | 20.0% | 13.3% | 0.8% | 1.0% |
| F1 Score(%)    | 90.2% | 31.0% | 88.4% | 30.0% | 18.7% |
| Precision(%)   | 87.7% | 61.9% | 83.7% | 62.0% | 62.0% |
|                |      |     |     |     |      |

Sleep EDF, Sleep EDFx, and DREAMS Subjects datasets by accuracy and kappa pair of (93.48, 0.898), (93.14, 0.890) and (83.50, 0.775), respectively. We can conclude from the obtained results, the algorithm is capable of performing on different datasets and has good generalizability.

IV. DISCUSSION AND FUTURE WORK

In this study, a novel algorithm is proposed for sleep staging which includes a simple and effective framework for feature extraction and a 2-stage learning strategy for classification that achieves robust and reliable results, and improves sleep staging results in comparison to the baseline methods. Now we discuss the important details of the automatic sleep staging problem and the significant improvements this study contributed to.

Finding the parameters of a CNN can be interpreted as an optimization problem that in each update of the parameters, loss value decreases, accuracy increases, and the model moves towards the optimal point. However, the network should be tested on different data than the train data and it is favorable that the model does not perfectly learn the train data to prevent overfitting. Therefore, regularization methods like the dropout layer can be used to prevent the model from overfitting.

Regularization methods by adding some error to the problem can improve results on test data but escalate the overshoot around the optimal point which can be seen in Fig. 4, 1-stage learning in loss curves. In this setup, the results are not reproducible and less reliable. One possible solution is to use a lower learning rate for the optimizer which can solve the problem to some extent but prolong the training process. However, finding the best learning rate has its complications. The proposed 2-stage learning is an alternative solution to overcome the mentioned problems and achieve improvements in consistency and accuracy of the model without any additional cost. Furthermore, in the second stage of the 2-stage learning, trainable parameters of the convolutional layers are frozen hence the computational cost and overall training time of the model are reduced.

The CNN structure was designed based on experimental investigation. Searching through different CNN architectures to find the model with the best performance for a specific application can be a cumbersome task. However, in our study, we tried to use a simple structure with good results and improved the model’s performance by two-stage learning. For choosing the best number of CNN streams and EEG subepochs, 4 cases are considered, including 1-stream to 4-stream CNN. Fig. 5 shows the accuracies for the different cases for the last 50 epochs of the training. It can be seen that the 2-stream CNN shows better performance in comparison to the other networks.

Despite the simple nature of the hand-crafted features in this study, they are capable of reflecting distinctive characteristics of the EEG signal for various sleep stages and improving sleep stage classification results. In the following, we discuss the provided information by features in detail. Fig. 6 illustrates the extracted features from an EEG subepoch which provides useful information about each sleep stage and its characteristics. The vertical axis is the logarithm of the calculated standard deviation of each EEG subband related to the 1 Hz frequency range. Each sleep stage curve was obtained by averaging over feature vectors of all of the epochs that scored as a specific sleep stage. This plot can be interpreted as an indicator of the frequency activity of each sleep stage signal in various frequency ranges. In the wake stage, theta waves are compressed and alpha waves have dominant activity. The broadband activity of the wake stage stands out in comparison to the other sleep stages. The unique shape of the wake curve makes it easy for the CNN to recognize it. However, as reported in confusion matrices in Table 5, Table 6 and Table 7 stage N1 has the lowest recognition rate among sleep stages and can be falsely
TABLE 8. Overall performance comparison of various methods on Sleep EDF, Sleep EDFx, and DREAMS Subjects datasets. The highest accuracy and kappa values are shown in bold.

| Dataset | Study | Year | Channel | No. of epochs | Accuracy(%) | kappa |
|---------|-------|------|---------|---------------|-------------|-------|
| Sleep EDF | Zhu et al. [15] | 2014 | Pz-Oz | 14963 | 88.9 | 0.83 |
| | Hassan et al. [37] | 2015 | Pz-Oz | 15188 | 86.5 | 0.787 |
| | Hassan et al. [38] | 2016 | Pz-Oz | 15188 | 90.1 | 0.844 |
| | Hassan et al. [39] | 2016 | Pz-Oz | 15188 | 90.7 | 0.855 |
| | Hassan et al. [40] | 2016 | Pz-Oz | 15188 | 91.5 | 0.865 |
| | Sharma et al. [41] | 2017 | Pz-Oz | 15136 | 91.1 | 0.861 |
| | Rahman et al. [42] | 2018 | EOG | 15188 | 91.02 | - |
| | Seifpour et al. [43] | 2018 | Pz-Oz | 15186 | 91.8 | 0.870 |
| | Yildirim et al. [44] | 2019 | Fpz-Cz | 15188 | 90.83 | - |
| | Jiang et al. [45] | 2019 | Fpz-Cz | 15160 | 92.7 | 0.884 |
| | Zhou et al. [46] | 2020 | Pz-Oz | 15170 | 91.2 | 0.864 |
| | Liu et al. [47] | 2021 | Pz-Oz | 15170 | 91.9 | - |
| | Huang et al. [48] | 2022 | Pz-Oz | 15188 | 91.29 | - |
| | Proposed algorithm | | Fpz-Cz | 15188 | **93.48** | **0.898** |
| Sleep EDFx | Alickovic et al. [6] | 2018 | Pz-Oz | 34288 | 91.1 | 0.879 |
| | Rahman et al. [42] | 2018 | EOG | 54587 | 92.6 | - |
| | Yildirim et al. [44] | 2019 | Fpz-Cz | 127512 | 90.48 | - |
| | Jadhav et al. [49] | 2020 | Fpz-Cz | 62177 | 83.17 | - |
| | Cai et al. [50] | 2020 | Fpz-Cz | 37226 | 89.1 | 0.8294 |
| | Delimayanti et al. [51] | 2020 | Pz-Oz | 127,663 | 91.42 | - |
| | Li et al. [52] | 2022 | Fpz-Cz | 42049 | 84.14 | 0.759 |
| | Huang et al. [48] | 2022 | Pz-Oz | 29750 | 89.37 | - |
| | Yang et al. [53] | 2022 | Fpz-Cz | 42308 | 85.2 | 0.79 |
| | Proposed algorithm | | Fpz-Cz | 36963 | **93.14** | **0.890** |
| DREAMS | Hassan et al. [40] | 2016 | Cz-A1 | 20,257 | 79.44 | - |
| | Seifpour et al. [43] | 2018 | Cz-A1 | 20257 | 83.3 | 0.770 |
| | Shen et al. [54] | 2020 | Cz-A1 | 20265 | 79.90 | - |
| | Liu et al. [47] | 2021 | Cz-A1 | 20242 | 83.2 | 0.77 |
| | Yang et al. [53] | 2022 | Pz-Oz | 20242 | 83.50 | 0.775 |
| | Proposed algorithm | | Cz-A1 | 20242 | - | - |

recognized as wake, N1, or REM. In addition to the few epochs this stage contains, the N1 stage can be considered the transition between the Wake and N2 stages and can be mislabeled as either one. On the other hand, stage N1 has a similar frequency activity to the REM stage as can be seen from Fig. 6 which makes it difficult for the classifier to distinguish N1 from the REM stages.

From the confusion matrices, the N3 epochs are falsely classified as N2 stage with 10.5% 13.4% and 18.7% rates for Sleep EDF, Sleep EDFx, and DREAMS dataset, respectively. This is a well-known problem in sleep staging that happens because of the presence of sleep spindles in both N2 and N3 stages. It can be seen from Fig. 6, N2 and N3 curves are overlapped in the frequency range of 12 to 16 Hz where the sleep spindles are active. However, the N3 sleep stage has a dominant frequency activity in the delta frequency band which makes it easy to recognize by the classifier.

The amplitude and frequency of the EEG signals are two important characteristics that change over time and determine the nonstationarity of the EEG and it does not depend on a specific application like sleep staging. The proposed approach utilizes these two characteristics for further processing. Therefore, the method for extracting hand-crafted features is not designed for a specific task such as sleep staging but to obtain a detailed understanding of the EEG. Hence, it can be used in different tasks that depend on EEG signals such as brain-computer interface [55] (BCI), emotion recognition [56] and seizure detection [57]. On the other hand, 2-stage learning sets a novel point of view in training CNNs that helps the model learn faster and achieve robust
M. AbdollahPour et al.: A Two-Stage Learning Convolutional Neural Network for Sleep Stage Classification

V. CONCLUSION

This paper presents a computer-aided sleep staging algorithm using a single-channel EEG signal. The main idea is based on introducing a method to extract the most informative features from EEG signal and classify them in a reliable framework with minimum limitations. A novel framework was introduced for feature extraction to achieve useful representations of the sleep EEG. In this framework, the EEG signal was decomposed into its frequency subbands by a filterbank and standard deviation as a single feature was calculated for each subband. These features provide distinctive characteristics of EEG signals for different sleep stages. A 2-stream CNN was designed for the classification, and a 2-stage learning method was used to train the model. Using 2-stage learning improved the model’s performance in terms of accuracy, robustness, and reproducibility. The performance of the method was evaluated from different aspects on three benchmark datasets, and it outperformed state-of-the-art sleep staging methods. Moreover, the introduced framework for classification outperformed conventional classifiers and 2-stage learning showed great improvement in comparison to traditional CNN training. The proposed method has the potential for practical implementation, and the need for only one channel of EEG signal makes it possible to use it on carry-on devices at home without supervision.

REFERENCES

[1] D. Riemann, M. Berger, and U. Voderholzer, “Sleep and depression: Results from psychobiological studies: an overview,” Biological psychology, vol. 57, no. 1-3, pp. 67–103, 2001.
[2] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, B. V. Vaughn et al., “The aasm manual for the scoring of sleep and associated events,” Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine, vol. 176, p. 2012, 2012.
[3] B. J. Swihart, B. Caffo, K. Blando-Roche, and N. M. Punjabi, “Characterizing sleep structure using the hypnogram,” Journal of Clinical Sleep Medicine, vol. 4, no. 4, pp. 349–355, 2008.
[4] H. Agnew Jr, W. B. Webb, and R. L. Williams, “The first night effect: an eeg study of sleep,” Psychophysiology, vol. 2, no. 3, pp. 263–266, 1966.
[5] N. Michielli, U. R. Acharya, and F. Molinari, “Cascaded lstm recurrent neural network for automated sleep stage classification using single-channel eeg signals,” Computers in biology and medicine, vol. 106, pp. 71–81, 2019.
[6] E. Alickovic and A. Subasi, “Ensemble svm method for automatic sleep stage classification,” IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 6, pp. 1258–1265, 2018.
[7] F. Andreotti, H. Phan, N. Cooray, C. Lo, M. T. Hu, and M. De Vos, “Multichannel sleep stage classification and transfer learning using convolutional neural networks,” in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018, pp. 171–174.
[8] R. Boostani, F. Karimzadeh, and M. Nami, “A comparative review on sleep stage classification methods in patients and healthy individuals,” Computer methods and programs in biomedicine, vol. 140, pp. 77–91, 2017.
[9] M. Abdollahpour, T. Y. Rezaei, A. Farzamnia, and S. Moshfeghi, “Sleep stage classification using dempster-shafer theory for classifier fusion,” in 2018 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET). IEEE, 2018, pp. 1–4.
[10] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, “Mixed neural network approach for temporal sleep stage classification,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 2, pp. 324–333, 2017.
[11] P. Memar and F. Faradji, “A novel multi-class eeg-based sleep stage classification system,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 1, pp. 84–95, 2017.
X. Li, L. Cui, S. Tao, J. Chen, X. Zhang, and G.-Q. Zhang, “Hyclass: a hybrid classifier for automatic sleep stage scoring,” IEEE journal of biomedical and health informatics, vol. 22, no. 2, pp. 375–385, 2017.

H. Phan, Q. Do, T.-L. Do, and D.-L. Vu, “Metric learning for automatic sleep stage classification,” in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2013, pp. 5025–5028.

S.-F. Liang, C.-E. Kao, Y.-H. Hu, and Y.-S. Cheng, “A rule-based automatic sleep staging method,” Journal of neuroscience methods, vol. 205, no. 1, pp. 169–176, 2012.

C. Zha, Y. Li, and H. Wen, “Analysis and classification of sleep stages based on differential visibility graphs from a single-channel eeg signal,” IEEE journal of biomedical and health informatics, vol. 18, no. 6, pp. 1813–1821, 2014.

S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

Y. Zhang, Z. Yang, K. Lan, X. Liu, Z. Zhang, P. Li, D. Cao, J. Zheng, and J. Pan, “Sleep stage classification using bi-dimensional lstm in wearable multi-sensor systems,” in IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2019, pp. 443–448.

H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, “Joint classification and prediction cnn framework for automatic sleep stage classification,” IEEE Transactions on Biomedical Engineering, vol. 66, no. 5, pp. 1285–1296, 2018.

J. Zhang, R. Yao, W. Ge, and J. Gao, “Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel eeg,” Computer methods and programs in biomedicine, vol. 183, p. 105089, 2020.

L. Jing, T. Parag, Z. Wu, Y. Tian, and H. Wang, “Videossl: Semi-supervised learning for video classification,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1110–1119.

T. Zha, W. Luo, and F. Yu, “Convolution-and attention-based neural network for automated sleep stage classification,” International journal of Environmental Research and Public Health, vol. 17, no. 11, p. 4152, 2020.

S. Chambon, M. N. Gaitler, P. J. Arnal, G. Wainrib, and A. Gramfort, “A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 4, pp. 758–769, 2018.

Z. Cui, X. Zheng, X. Shao, and L. Cui, “Automatic sleep stage classification based on convolutional neural network and fine-grained segments,” Complexity, vol. 2018, 2018.

F. Li, R. Yan, R. Mahini, L. Wei, Z. Wang, K. Mathiak, R. Liu, and F. Cong, “End-to-end sleep staging using convolutional neural network in raw single-channel eeg,” Biomedical Signal Processing and Control, vol. 63, pp. 102020, 2020.

Y.-D. Zhang, Z. Dong, X. Chen, W. Jia, S. Du, K. Muhammad, and S.-H. Wang, “Image-based fruit category classification by 13-layer deep convolutional neural network and data augmentation,” Multimedia Tools and Applications, vol. 78, no. 3, pp. 3613–3632, 2019.

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.

B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Obery, “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.

A. Rechtshaffen, “A manual for standardized terminology, techniques and classification and prediction cnn framework for automatic sleep stage classification using difference visibility graphs from a single-channel eeg signal,” IEEE journal of biomedical and health informatics, vol. 18, no. 6, pp. 1813–1821, 2014.

J. A. Onton, D. Y. Kang, and T. P. Coleman, “Visualization of whole-night sleep eeg from 2-channel mobile recording device reveals distinct deep sleep stages with differential electrodermal activity,” Frontiers in human neuroscience, vol. 10, p. 605, 2016.

Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, “Automatic sleep stage classification using energy features of eeg signals,” Neurocomputing, vol. 104, pp. 105–114, 2013.

R. Keshari, M. Vatsa, R. Singh, and A. Noore, “Learning structure and strength of cnn filters for small sample size training,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9349–9358.

M. Abdollahpour, T. Y. Rezai, A. Farzamnia, and I. Saad, “Transfer learning convolutional neural network for sleep stage classification using two-stage data fusion framework,” IEEE Access, vol. 8, pp. 180618–180632, 2020.

D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

J. Cohen, “A coefficient of agreement for nominal scales,” Educational and psychological measurement, vol. 20, no. 1, pp. 37–46, 1960.

A. R. Hassan, S. K. Bashar, and M. I. H. Bhiuian, “On the classification of sleep stages by means of statistical and spectral features from single channel electroencephalogram,” in 2015 International conference on advances in computing, communications and informatics (ICACCI). IEEE, 2015, pp. 2238–2243.

A. R. Hassan and M. I. H. Bhiuian, “Automatic sleep scoring using statistical features in the emd domain and ensemble methods,” Biocybernetics and Biomedical Engineering, vol. 36, no. 1, pp. 248–255, 2016.

A. R. Hassan and M. I. H. Bhiuian, “Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating,” Biomedical Signal Processing and Control, vol. 24, pp. 1–10, 2016.

A. R. Hassan and M. I. H. Bhiuian, “A decision support system for automatic sleep staging from eeg signals using tunable q-factor wavelet transform and spectral features,” Journal of neuroscience methods, vol. 297, pp. 107–118, 2018.

R. Sharma, R. B. Pachori, and A. Upadhyay, “Automatic sleep stages classification based on iterative filtering of electroencephalogram signals,” Neural Computing and Applications, vol. 28, no. 10, pp. 2959–2978, 2017.

M. M. Rahman, M. I. H. Bhiuian, and A. R. Hassan, “Sleep stage classification using single-channel eeg,” Computers in biology and medicine, vol. 102, pp. 211–220, 2018.

S. Seifpour, H. Nikzahar, M. Mikaeili, and A. M. Namrabad, “A new automatic sleep staging system based on statistical behavior of local extrema using single channel eeg signal,” Expert Systems with Applications, vol. 104, pp. 277–293, 2018.

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, p. 599, 2019.

D. Jiang, Y.-n. Lu, M. Yu, and W. Yuanhua, “Robust sleep stage classification with single-channel eeg signals using multimodal decomposition and hmm-based refinement,” Expert Systems with Applications, vol. 121, pp. 188–203, 2019.

J. Zhou, G. Wang, J. Liu, D. Wu, W. Xu, Z. Wang, J. Ye, M. Xia, Y. Hu, and Y. Tian, “Automatic sleep stage classification with single channel eeg signal based on two-layer stacked ensemble model,” IEEE Access, vol. 8, pp. 57263–57297, 2020.

C. Liu, B. Tan, M. Fu, J. Li, J. Wang, F. Hou, and A. Yang, “Automatic sleep staging with a single-channel eeg based on ensemble empirical mode decomposition,” Physica A: Statistical Mechanics and its Applications, vol. 567, p. 125685, 2021.

Z. Huang and B. W.-K. Ling, “Joint ensemble empirical mode decomposition and tunable q factor wavelet transform based sleep stage classifications,” Biomedical Signal Processing and Control, vol. 77, p. 103760, 2022.

P. Jadhav, G. Raigurur, D. Datta, and S. Mukhopadhyay, “Automatic sleep stage classification using time–frequency images of eeg and transfer learning using convolution neural network,” Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 494–504, 2020.

Q. Cai, Z. Gao, J. An, S. Gao, and C. Grebogi, “A graph-temporal fused dual-input convolutional neural network for detecting sleep stages from eeg signals,” IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 68, no. 2, pp. 777–781, 2020.

M. K. Delimayanti, B. Purnama, N. G. Nguyen, M. Faisal, K. R. Mahmodud, F. Indriani, M. Kubo, and K. Satou, “Classification of brainwaves
for sleep stages by high-dimensional fft features from eeg signals," *Applied Sciences*, vol. 10, no. 5, p. 1797, 2020.

[52] T. Li, B. Zhang, H. Lv, S. Hu, Z. Xu, and Y. Tuergong, “Cattsleepnet: Automatic end-to-end sleep staging using attention-based deep neural networks on single-channel eeg,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 9, p. 5199, 2022.

[53] B. Yang, W. Wu, Y. Liu, and H. Liu, “A novel sleep stage contextual refinement algorithm leveraging conditional random fields,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

[54] H. Shen, F. Ran, M. Xu, A. Guez, A. Li, and A. Guo, “An automatic sleep stage classification algorithm using improved model based essence features,” *Sensors*, vol. 20, no. 17, p. 4677, 2020.

[55] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, “Optimizing the channel selection and classification accuracy in eeg-based bci,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1865–1873, 2011.

[56] R. Jenke, A. Peer, and M. Buss, “Feature extraction and selection for emotion recognition from eeg,” *IEEE Transactions on Affective computing*, vol. 5, no. 3, pp. 327–339, 2014.

[57] L. Wang, W. Xue, Y. Li, M. Luo, J. Huang, W. Cui, and C. Huang, “Automatic epileptic seizure detection in eeg signals using multi-domain feature extraction and nonlinear analysis,” *Entropy*, vol. 19, no. 6, p. 222, 2017.

[58] A. Malhotra, I. Ayappa, N. Ayas, N. Collop, D. Kirsch, N. Mcardle, R. Mehra, A. I. Pack, N. Punjabi, D. P. White et al., “Metrics of sleep apnea severity: beyond the apnea-hypopnea index,” *Sleep*, vol. 44, no. 7, p. zsab030, 2021.

***