Automatic Sleep Scoring Using Intrinsic Mode Based on Interpretable Deep Neural Networks

JAEWOO BAEK[1], **CHOONGSEOP LEE**[1], **HYUNSOO YU**[1], **SUWHAN BAEK**[1], **SEOKMIN LEE**[2], **SEUNGMIN LEE**[3], **AND CHEOLSOO PARK**[1], **(Member, IEEE)**

[1]Department of Computer Engineering, Kwangwoon University, Seoul 01897, Republic of Korea
[2]School of Cybersecurity, Korea University, Seoul 02841, South Korea
[3]Department of Electrical Engineering, Kookmin University, Seoul 02707, Republic of Korea

Corresponding authors: Seungmin Lee (smlee27@kookmin.ac.kr) and Cheolsoo Park (parkcheolsoo@kw.ac.kr)

ABSTRACT Sleep experts manually label sleep stages via polysomnography (PSG) to diagnose sleep disorders. However, this process is time-consuming, requires a lot of labor from sleep experts, and makes the participants uncomfortable with the attachment of multiple sensors. Thus, automatic sleep scoring methods are essential for practical sleep monitoring in our daily lives. In this study, we propose an automatic sleep scoring model based on intrinsic oscillations in a single channel electroencephalogram (EEG) signal. We applied noise assisted bivariate empirical mode decomposition (NA-BEMD) to extract the intrinsic mode components and an attention mechanism in deep neural networks to provide weights to the components depending on their significance to sleep scoring. In particular, through the attention mechanism, we found an interpretable model by examining the oscillations that correspond to specific sleep stages. Therefore, we analyzed which frequency components are more weighted to a sleep stage than the others, when the model classifies sleep stages, and, as a result, confirmed that the model assigns convincing weights to the frequency components for each sleep stage. Additionally, the model consists of a one-dimensional convolutional neural network (1D-CNN) to extract features of an epoch and bidirectional long short-term memory (Bi-LSTM) to learn the sequential information of the consecutive epochs. We evaluated proposed model using Fpz-Cz, Pz-Oz, and F3-M2 channel EEG from three different public datasets (Sleep-EDF-2013, Sleep-EDF-2018, WSC) and demonstrated that our model yielded the best overall accuracy (Fpz-Cz: 86.22%-82.67%, Pz-Oz: 83.63%-80.15%, F3-M2: 84.20%) and macro F1-score (Fpz-Cz: 80.79%-76.90%, Pz-Oz: 76.89%-72.98%, F3-M2: 74.88%) compared with the state-of-the-art sleep scoring algorithms using single channel EEG. As a benchmark test, FIR bandpass filters were compared, and it was confirmed that NA-BEMD was superior to the traditional filters in all experiments, demonstrating that the proposed model is interpretable and a state-of-the-art sleep scoring algorithm.

INDEX TERMS Electroencephalogram (EEG), automatic sleep scoring, deep neural networks, attention mechanism, bivariate empirical mode decomposition.

I. INTRODUCTION

Sleep is a crucial factor to maintain a healthy life for human beings. Deterioration of sleep quality causes various sleep disorders such as insomnia, sleep apnea, and parasomnia [1]. Numerous people suffer from sleep disorders, which are diagnosed by sleep experts who conduct a multi-parametric test called polysomnography (PSG) [2]. The classification

of multiple sleep stages, called sleep scoring or sleep stage scoring, is carried out through PSG. Sleep experts determine sleep quality by analyzing the distribution of each stage across all night sleep stages. PSG monitoring measures physiological signals including electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EOG), electromyography (EMG). Based on the study of Rechtschaffen and Kales (R & K) [3] or the American Academy of Sleep Medicine (AASM) [4], sleep stages are decided over 30-second intervals called epochs, by sleep

The associate editor coordinating the review of this manuscript and approving it for publication was Kin Fong Lei.

experts. In the R&K standard, PSG signals are classified into six stages: wakefulness (Wake), non-rapid eye movement (NREM) stages (N1-N4), and rapid eye movement (REM). The AASM standard combines the N3 and N4 stages into slow-wave sleep (SWS) stages or N3 stages. In this study, we use the AASM standard to classify the five sleep stages.

Manual sleep scoring by sleep experts is time-consuming, requires a lot of labor, and makes participants feel uncomfortable due to a number of sensors attached to them during sleep. In addition, different sleep scoring results are produced by different sleep experts, even during re-scoring by the same expert [5]. To solve these issues associated with manual sleep scoring, automatic sleep scoring algorithms have been developed using a single-channel EEG. In contrast to the manual approach, automatic sleep scoring is consistent and reduces the scoring time, labor, and cost [6].

Several state-of-the-art machine learning and deep learning algorithms for the analysis of biosignals have been suggested for automatic health diagnosis and monitoring systems. In [7],the authors proposed the prediction of stroke using the random forest algorithm based on ECG signals. In addition, a health monitoring system for the stroke prognostics using SVM is proposed by [8]. In [9], convolutional neural network (CNN) [10] was proposed for the seperation of the motor imaginary responses using EEG signals. Moreover, Senyurek *et al*. [11] analyzed respiratory signals for the puff detection based on CNN and long short term memory (LSTM) [12].

Early studies on automatic sleep scoring extracted time-frequency features from an epoch of a single channel EEG and employed machine learning algorithms to classify the sleep stages. For example, Oropesa *et al*. [13] used wavelet packet transformation (WPT) to provide localized time-frequency components and estimate the sleep stages using an artificial neural network (ANN). A continuous wavelet transform (CWT) was utilized by Fraiwan *et al*. [14] for their time-frequency component features obtained from EEG signals, which were classified using linear discriminant analysis (LDA). Furthermore, Fraiwan *et al*. [15] applied Choi–Williams distribution (CWD), Hilbert–Huang transform (HHT), and the CWT for instantaneous feature extraction, which was classified using a random forest algorithm. In addition, Sharma *et al*. [16] decomposed EEG signals into amplitude and frequency-modulated components using an iterative filtering method to extract the sleep stage-related features and applied naïve Bayes, k-nearest neighbor, multi-layer perceptron, decision tree, and random forest algorithms for classification.

Recently, deep learning approaches have been applied for automatic sleep scoring algorithms with or without traditional signal processing, which exhibit superior performance than conventional methods. In particular, the architectures of CNN and recurrent neural network (RNN) [17] are mainly applied for sleep scoring algorithms. Because CNN is well known to extract accurate feature information from input data, they can well recognize the characteristics of the

EEG amplitude or frequency components. Additionally, RNN has a memory cell architecture to utilize the temporal features of the EEG signals across consecutive epochs to estimate sleep stages. For example, Yildirim *et al.* [18] constructed a 16-layer one-dimensional CNN (1D-CNN) to extract EEG features and classified six sleep stages using fully connected layers and a softmax function. In addition, Zhu *et al.* [19] divided an epoch into 29 multiple windows to effectively capture sleep spindles or k-complex features and applied 1D-CNN and a self-attention mechanism for each window. In another study, 1D-CNN and Bi-LSTM were combined to estimate five class sleep stages [20], where 1D-CNN was pretrained using oversampled data to solve the class imbalance problem. There are other approaches that combine frequency analysis and deep learning algorithms. For example, Phan *et al.*. [21] employed Short Time Fourier transform (STFT) to produce time-frequency representations of a single channel EEG and then conducted sleep scoring using 1D-CNN. Michielli *et al.*. [22] extracted 55 features of EEG in the time and frequency domain using various statistical parameters, an infinite impulse response (IIR) bandpass filter, and a power spectrum, which were fed into the LSTM to classify four class sleep stages.

Currently, the state-of-the-art deep learning models for the automatic sleep scoring use the frequency components extracted through Fourier-based decomposition algorithm such as Fast Fourier Transform (FFT) or only raw EEG signals as their input [23]–[26]. The Fourier-based decomposition algorithms could not deal with reflect the non-linearity and non-stationarity of EEG signals [27]. If the raw EEG signal is fed into the deep learning model, the combined characteristics of the frequency components could be learned. However, it would be difficult for the model to train the features of each frequency components. In addition, it is crucial to look into the frequency components separately since several studies have already established specific components corresponding to a certain sleep stage [28]. For this issue, we propose a deep learning model with a data-driven frequency decomposition algorithm and the attention mechanism for the automatic sleep scoring. The main contributions of this paper are as follows.

- We apply a noise assisted bivariate empirical mode decomposition (NA-BEMD) to extract frequency components with its dyadic filter bank property. As a data-driven model, NA-BEMD can extract intrinsic oscillations in a signal, which can be accurately decomposed the sleep stage-related components in EEG signals [29].
- We design a novel deep learning architecture consisting of 1D-CNN and Bi-LSTM to extract features from each epoch of an EEG signal. The 1D-CNN with residual connections extract feature of an epoch, and Bi-LSTM learns the sleep stage transition rule using features of epochs generated by the 1D-CNN.
- We employ the attention mechanism to choose important frequency components for each sleep stage, producing weights to the input nodes of the networks based

TABLE 1. Epoch numbers corresponding to sleep stages in Sleep-EDF and Wisconsin sleep cohort dataset.

Dataset	Wake	N1	N2	N3	REM	Total	Participants
Sleep-EDF-2013	8285	2804	17799	5703	7717	42308	20
Sleep-EDF-2018	65628	21379	68136	12857	25547	193547	78
WSC	21060	8261	54218	7507	13067	104113	116

on their significance [30]. The frequency components decomposed through NA-BEMD are the input of the deep neural networks, and their attention weights are decided based on their contributions to the classification of each sleep stage.

- The weight information through the attention mechanism could make us understand how the deep neural network architecture processes the input data to obtain the final classification results rather than a black-box function.

- We demonstrate that the proposed model could classify 5 sleep stages using three different datasets, recorded in different conditions, without modifying the model hyperparameters and architecture, inferring the accomplishment of the generalized sleep scoring model.

This paper is structured as follows. Section II illustrates the public datasets used in the experiment, the proposed model architecture, and the evaluation method of the model in detail. In Section III, the performance of the generalized model is demonstrated using the evaluation process and compared with the existing works. In Section IV, the main contribution is discussed in more depth based on the experimental results, and the last section concludes the paper and describes the future research directions.

II. MATERIALS AND METHODS

A. DATASETS

We employed three different channels of EEG from three public datasets to train and evaluate the performance of the proposed method. There are two Sleep-EDF dataset (2013 and 2018 version) [31] and Wisconsin sleep cohort dataset (WSC) [32]. In this study, a single-channel of EEG signal is analyzed considering the development of convenient sleep monitoring device with minimum number of wires. The distributions of epochs from the 5 sleep stages and the number of participants for each dataset are described in Table 1. As shown in Table 1, the numbers of epochs among all the sleep stages are unbalanced. In addition, WSC has the highest number of participants among the datasets, and Sleep-EDF-2018 has the largest number of epochs among the datasets.

1) SLEEP-EDF DATASET

The Sleep-EDF dataset is often used in many automatic sleep scoring studies [33]. There are three PSG records, consisting of EEG signals with configuration of Fpz-Cz and Pz-Oz channel, EOG signals, and chin EMG signals, recorded at a sampling rate of 100Hz. In addition, it contains hypnogram information with annotation of six sleep stages labeled based on R&K standard. We combined the N3 and N4 stages into the N3 stage to satisfy the latest AASM standard [4]
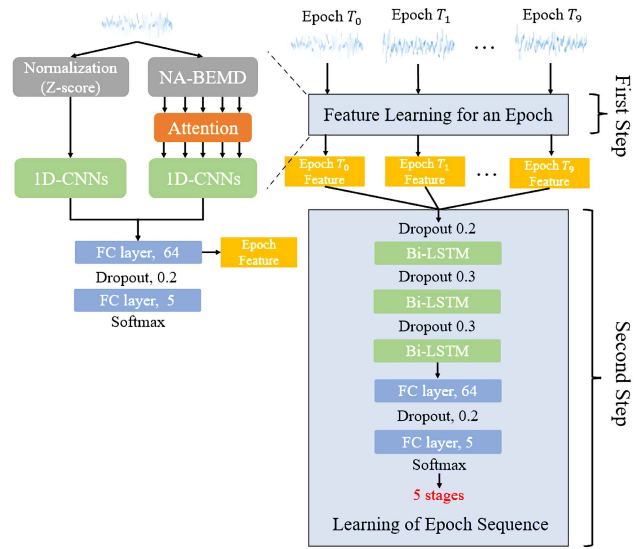

FIGURE 1. The architecture of our model. Model is trained for two step. Feature learning for an epoch, first training step, includes 1D-CNN layers and a attention layer. Learning of epoch sequence, second training step, includes Bi-LSTM layers. Each layer containing trainable parameters is optimized during two-step training.

with five sleep stages classification. There are two different participant groups in the Sleep-EDF dataset, that is, Sleep Cassette Study(SC) and Sleep Telemetry Study (ST) groups. SC group records were obtained in the study about the age effects on the sleep of the healthy participants aged between 25 and 101 years old, without any sleep-related medication. On the other hand, ST group was recorded in the study of the temazepam effects on the sleep of the insomnia participants. We only utilized records of healthy participants from SC group in the 2013, 2018 versions, which have records from 20 and 78 participants, respectively, on the Fpz-Cz and Pz-Oz channel configurations.

2) WISCONSIN SLEEP COHORT DATASET

Wisconsin Sleep Cohort (WSC) dataset includes 2570 PSG records from 116 participants of various races with various physiological signals such as EOG, EEG, EMG, ECG, and SpO2 [32]. EEG was recorded on the configurations of F3-M2, Fz-M2, Cz-M2, C3-M2, Pz-M2, O1-M2 channels at a sampling rate of 200 Hz, and F3-M2 channel data was utilized for our benchmark test. In WSC dataset, the participants have a maximum of 5 records, obtain during 5 different visits of the participants (visit1-visit5). In this study, we employed the dataset from visit4 to visit5, which are the most recent data group in WSC dataset and they had become sufficiently familiar with the laboratory environments.

B. OVERVIEW OF THE MODEL

Fig 1. illustrates the overall architecture of the proposed model for the automatic sleep scoring. After the segmentation of the EEG signal into a 30-second epoch, the raw signal in an epoch is decomposed into 5 components using NA-BEMD, corresponding to the important frequency components to the sleep scoring, delta (0.5-4Hz), theta (4-8Hz), alpha (8-13Hz),

beta (13-30Hz), and gamma (30Hz-) band components [34]. Also, at the same time, the epochs are z-score normalized. The five decomposed components are fed into the input layer of the proposed networks together with the normalized EEG signal as Z-scores. They are weighted through an attention layer, and the features of the 5 weighted frequency components and the normalized EEG signal are extracted respectively through the 1D-CNN. The extracted features of 10 epochs are considered as one sequence, and Bi-LSTM learns the successive epoch patterns among the 10 epochs. Additionally, the model is trained in two-step to more train network of first step than network of second step, since network of first step is deeper and more complex than second step and cause overfitting issue faster. Therefore, the model is trained network of second step using pre-trained network of first step as backbone. Through this, the generalization performance of the model was improved with preventing the overfitting issue of network of first step.

C. MODEL DESCRIPTION

1) NOISE ASSISTED BIVARIATE EMPIRICAL MODE DECOMPOSITION

Empirical mode decomposition (EMD) is a data-driven frequency decomposition algorithm in an adaptive way, which decomposes signals without any basis function [35], and thus it could well extract the intrinsic oscillations from the nonlinear EEG signals as demonstrated by Park et al.. [36] In contrast with the wavelet and Fourier analysis, EMD has few restrictions on a signal and no fixed basis function [37]. A raw signal $x(t)$ is sequentially decomposed into intrinsic mode functions (IMFs), $c_k(t)$ and residue $r_n(t)$, through the iterations of the sifting process (see Eq 1).

$$x(t) = \sum_{k=1}^{n} c_k(t) + r_n(t) \qquad (1)$$

During the sifting process, upper envelope $e_u(t)$ and lower envelope $e_l(t)$ are identified with interpolating all local maxima and minima, and compute the local mean $m(t)$ of these two envelopes. After this, $m(t)$ is repeatedly subtracted from the input data. If the residue $r_k(t)$ after the subtraction satisfies the two stopping criteria, $c_k(t)$ is assigned with $r_k(t)$. The two stopping criteria are defined as follows: first, the mean of the upper and lower envelopes is close to zero, and second, the number of extreme values and zero crossing points should be the same or only differ by one. Algorithm 1 describes the details of EMD.

BEMD was originally developed for complex-valued time series data [38]. Unlike the conventional univariate EMD, BEMD can process bivariate data in the complex-valued form and decompose it to complex-valued IMFs, which have real and imaginary parts for the bivariate data. The procedure of the BEMD algorithm is elaborated in Algorithm2.

Since the EEG signal is non-stationary and non-linear time series data, the EMD method could be suitable to decompose EEG signals due to its data driven process [35]. However, the distribution of the frequency components in each IMFs

---

Algorithm 1 Procedure of Empirical Mode Decomposition

1: Let $\tilde{x}(t) = x(t)$
2: Find all local maximum and minimum values of $\tilde{x}(t)$
3: Identify upper envelope $e_u(t)$ and lower envelope $e_l(t)$, through a cubic spline interpolation
4: $m(t) = (e_u(t) + e_l(t))/2$
5: $r_k(t) = \tilde{x}(t) - m(t)$
6: If the condition of $r_k(t)$ meet the stopping criterion, assign $c_k(t)$ with $r_k(t)$, otherwise $\tilde{x}(t) = r_k(t)$ and iterate step 2-5
7: $\tilde{x}(t) = \tilde{x}(t) - c_k(t)$
8: Iterate step 2-7, until $r_k(t)$ has no more oscillations

Algorithm 2 Procedure of Bivariate Empirical Mode Decomposition

1: Set the direction $d_n = 2n\pi/D \ (1 \leq n \leq D), D = 2$
2: Project the complex-valued signal $x(t)$ on a direction d_n: $p_{d_n}(t) = Re(e^{-j2n\pi/D}x(t))$
3: Find the local maxima $\{(t_i^n, p_i^n)\}$ of $p_{d_n}(t)$
4: Interpolate the set $\{(t_i^n, e^{-i2n\pi/D}p_i^n)\}$, building a tube in time, real and imaginary axes
5: Obtain the tangent g'_{d_n} on a tube in the direction of d_n
6: Iterate step 2-5 twice with n = 1, 2
7: Average all tangents: $m(t) = 2(g'_{d_1} + g'_{d_2})/D$
8: From this step, it is the same as sifting processing of the conventional EMD.

---

decomposed using EMD is not uniform. For example, the range of the frequency components in the first IMF of one EEG signal would be different from the frequency range of the other EEG signal. Thus, even if an EEG training dataset including a specific order of IMF is learned by a classification model, the model might not classify the other testing dataset due to the inconsistent frequency distribution of the IMFs. In order to solve the above issue, we apply to a noise-assisted BEMD using two channels which are a raw single channel EEG and a white Gaussian noise signal with 0.1 standard deviations of the EEG. Since the IMFs of a white Gaussian noise have a dyadic filter bank property, it could also derive the similar dyadic frequency components in the EEG IMFs of the other channel, which could solve the mode mixing problem within the decomposed IMFs [39].

In this study, we applied NA-BEMD to the epochs to extract six IMFs covering a frequency range from 0.5Hz to 50Hz. One example of the six IMFs from an epoch are displayed in Fig 2., and it shows that NA-BEMD produces the different oscillations in data-driven way considering the non-linearity and non-stationarity of the EEG signal. In addition, the averaged power spectra of all epochs' IMFs of the participant are shown in Fig 3., where the dyadic filter bank property can be noted. As the sifting process of NA-BEMD proceeds, slower oscillations are extracted than the earlier components. IMF1 contains high frequency noise, which could be found heuristically during an experiment
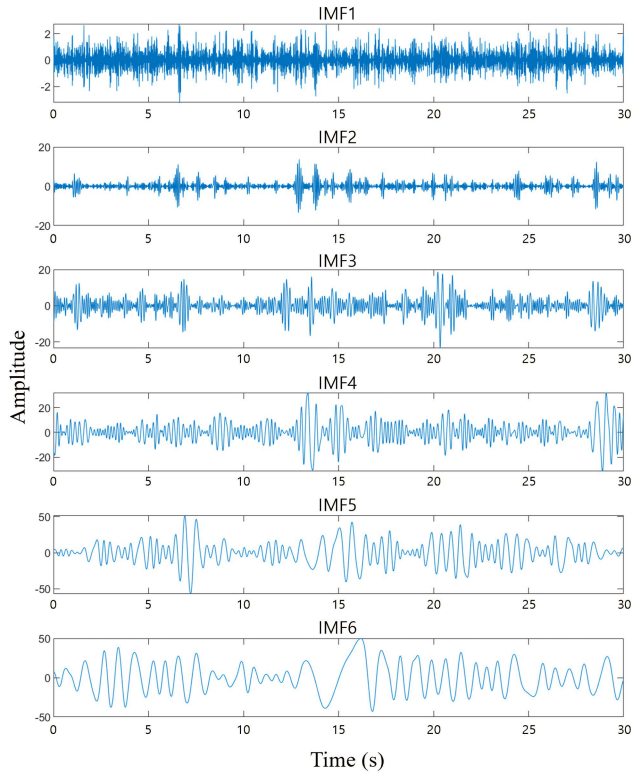
FIGURE 2. One example of six IMFs from an epoch of a participant, 'SC4001E0', in the Sleep-EDF dataset. IMF2 - IMF6 are fed into the Attention layer and IMF1 is discarded since high frequency noise.
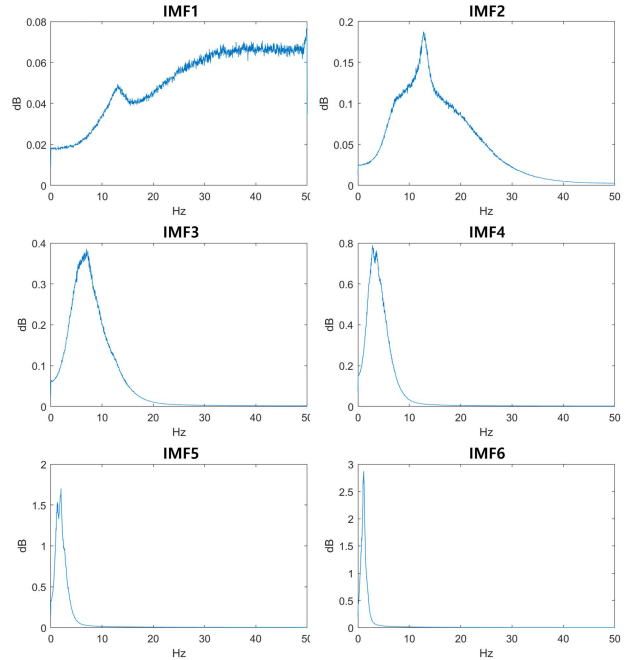

FIGURE 3. Averaged power spectra of the EEG IMFs from a participant, 'SC4001E0' in the Sleep-EDF dataset. Note the dyadic filterbank property of the IMFs decomposed using NA-BEMD.
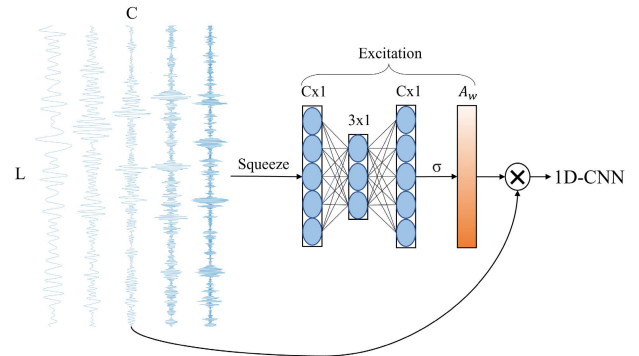
learning the sleep scoring model using all 6 IMFs and 5 IMFs discarding IMF1. The 5 IMFs without IMF1 improved the sleep scoring performance, and thus the proposed model learned the 5 components only.

2) ATTENTION MECHANISM

Fig 4. illustrate the structure of an attention layers. The attention layer generates attention weights and executes element-wise multiplication to the input data. This attention mechanism consists of squeeze and excitation operations [30]. In squeeze operation, the five IMFs are fed into the global average pooling (GAP) layer to compress the global spatial information into a descriptor. The excitation operation computes the dependencies among the components using fully connected layers and non-linear functions. The formulas for these operations are as follows in Eq 2-4.

$$G_c = GAP(x_c(t)) = \frac{1}{L} \sum_{t=1}^{L} x_c(t) \tag{2}$$

$$A_w = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 G_c)) \tag{3}$$

$$x_{atten}(t) = x_c(t) \odot rescale(A_w) \tag{4}$$

where δ and σ denote the Relu and sigmoid functions. $\mathbf{W}_1$ and $\mathbf{W}_2$ are trainable weight vector in the attention layer, $\mathbf{W}_1 \in \mathbb{R}^{3 \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times 3}$. The attention weights, A_w, extracted through the sigmoid function are multiplied by the IMFs. Through the attention layers, the performance is

FIGURE 4. Attention layer before the 1D-CNN layers. C denotes the number of samples in a 30-second epoch. Squeeze operation is global averaging pooling (GAP) and blue circles represent fully connected layers. A_w is attention weight that is the output of the sigmoid function.

improved by focusing on the important IMFs for classifying the sleep stage. An IMF with a large attention weight could contain significant features for the model task. Since the outputs of the sigmoid function have values between 0 and 1, they are scaled corresponding to the significance of the components [30].

3) 1D-CNN FOR AN EPOCH's FEATURE LEARNING

We use 1D-CNN with various filter sizes in all the convolution layers to extract features from the decomposed 5 IMFs and raw EEG signal. A large sized filter would detect large patterns well, while a small sized filter would recognize small patterns [40]. In detail, there are 13 convolution layers with various filter sizes instead of the fixed size, 6 max-pooling
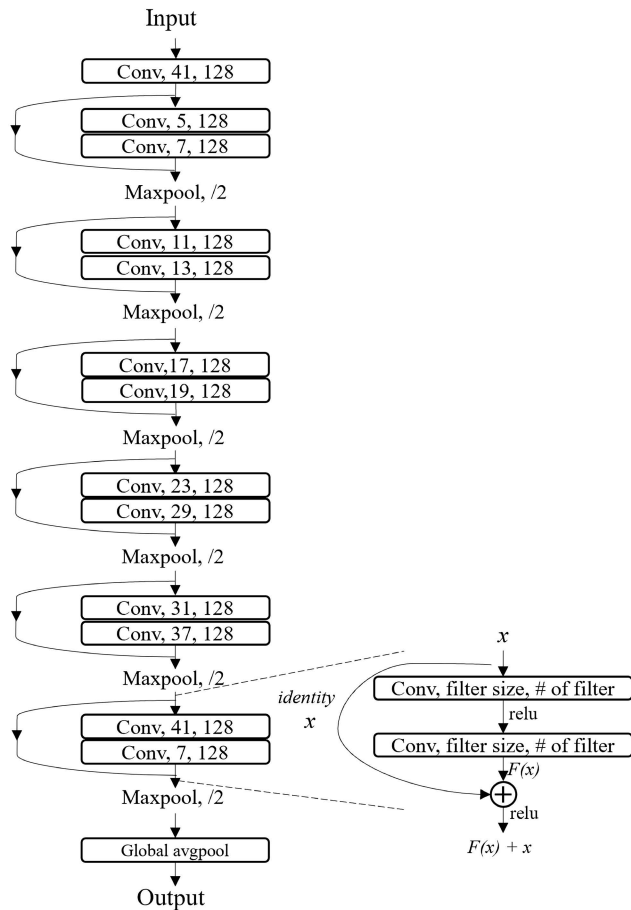
FIGURE 6. Architecture of Bi-LSTM consist of forward and backward direction LSTM. The EEG features from x_0 to x_9 extracted from 10 epochs are fed into the first Bi-LSTM layer, which could learn the sequential patterns among the 10 epochs.

FIGURE 5. The architecture of 1D-CNN with the residual connections. The residual connection is the process of adding the output $F(x)$ into the identity x of the trainable layers. The stride of all convolution layers and max-pooling layers is 1 and 2 respectively. This architecture is applied to both IMF components and raw EEG signals.

layers, and global average pooling layer in the 1D-CNN, shown in Fig 5.. Each convolution layer has the ReLU, which is a non-linear activation function. In addition, we apply the residual connection to solve the vanishing gradient problem and to ease the training of the network by adding identity x to the output $F(x)$ [41]. Therefore, this method allows the CNN layers to be deeper and improve the performance with the increased layer depth. [42] In addition, the residual connections are relatively easier to be optimized, compared with vanilla CNN. The features extracted from the IMF components and the normalized raw signal are concatenated and processed through two fully connected layers, and finally, five sleep stages are classified using the softmax function shown in Fig 1.. The extracted features from 10 epochs are transferred to the next Bi-LSTM layers.

4) BI-LSTM FOR LEARNING OF EPOCH SEQUENCE

In general, successive sleep stages have regular patterns [43]. For example, if the participant is in the wake stage, then the next stage is likely to be the wake or N1 stage. In addition, N1 and N2 stages have similar patterns of the brain waves,
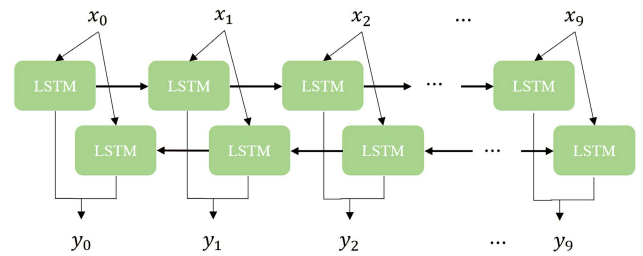
that is both stages have low amplitude and mixed frequency properties in EEG signal. However, if sleep spindles or k-complexes appear during a few minutes, all epochs in that duration are scored as N2 stages. Therefore, in order to decide a correct sleep stage between N1 and N2 or between Wake and N1, a learning structure based on the consecutive epochs is required. To learn this rule, the proposed model considers 10 epochs as one set and uses three Bi-LSTM layers that could memorize information of the previous or subsequent epochs [44]. Bi-LSTM layers can process sequences in forward and backward direction with merging two LSTMs. The input sequence x_t is a feature vector of 10 epochs extracted using 1D-CNN structure and t denotes the time index in Bi-LSTM. This structure is formulated as follows:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(x_t, h_{t-1}) \tag{5}$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(x_t, h_{t-1}) \tag{6}$$

$$y_t = \overrightarrow{h_t} \,||\, \overleftarrow{h_t} \tag{7}$$

where $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are hidden states of the forward and backward networks. Finally, the output y_t is made by concatenating both hidden states. Thus, it could learn past and future information. Fig 6. illustrates the Bi-LSTM layer architecture. The outputs of Bi-LSTM layer from y_0 to y_9 are fed into the next Bi-LSTM layer, and these Bi-LSTM layers are stacked in three layers in the model.

5) TRAINING DETAIL

In this study, the hyperparameters of our model were designed experimentally through trial and error, where several configurations were tried with twenty-fold cross validations using the Sleep-EDF-2013 dataset. We adjusted experimentally hyperparameters of 1D-CNN and Bi-LSTM, the number of convolution layers and fully connected layers, epoch sequence length, learning epochs, learning rate, and batch size. We used the Adam optimizer with a learning rate of 0.0001 and categorical cross-entropy as the loss function. The batch size is 80 when batch feds to network of feature learning for an epoch. 10 features of epochs are bundled into one set and 8 sets fed to network of learning of epoch sequence. In Fig 1., the output shape of the first fully connected layer of 1D-CNN is (8,10,64), where the time step and the number of attributes of Bi-LSTM are 10 and 64.

In addition, each learning epoch of an EEG epoch feature extraction learning and EEG epochs sequence learning is 8 and 50. Furthermore, the weight initializer in the model is GlorotUniform and its random seed is 777.

D. EXPERIMENTAL SETUP

1) EVALUATION METHOD

Ten and twenty-fold cross validations were conducted to validate the performance of the trained model [45]. In addition, the participants' test data were completely blinded to the training dataset for a fair inter-participant test. It should except epochs of the test participants from the training dataset. Therefore, because there are 20 participants in the Sleep-EDF-2013 dataset, twenty-fold cross validation was conducted, and ten-fold cross validation was performed for the Sleep-EDF-2018 dataset and WSC dataset. This inter-participant cross validation is practical to implement real world applications.

2) ATTENTION WEIGHT ANALYSIS METHOD

When the proposed model classifies the sleep stages, the attention weights are calculated and averaged during the cross validation to analyze which frequency components are weighted by the attention layer in a specific stage. In addition, One-tailed t-test is also conducted to determine whether the difference of the attention weights assigned to the frequency components in each sleep stage is significant.

3) COMPARISON WITH FIR FILTER

To evaluate the performance of the NA-BEMD preprocessing, the Fourier based FIR bandpass filter (BF) was compared as a benchmark test. EEG signals are decomposed into several sub-waves, beta (13-30Hz), alpha (8-13Hz), theta (4-8Hz), delta (0.5-4Hz) waves, which were often used for sleep scoring [46]. Corresponding to these frequency bands, IMF2, IMF3, IMF4, IMF5 and IMF6 decomposed using NA-BEMD were compared with fixing the model architecture.

4) PERFORMANCE METRICS

To evaluate the performance of the proposed model, three performance metrics— recall (*RE*), precision (*PRE*), and F1-score (*F1*)— were calculated. True positive (*TP*), true negative (*TN*), false negative (*FN*), and false positive (*FP*) were used to calculate the performance metrics as follows:

$$PRE = \frac{TP}{TP + FP} \quad (8)$$

$$RE = \frac{TP}{TP + FN} \quad (9)$$

$$F1 = \frac{2 \times PRE \times RE}{PRE + RE} \quad (10)$$

In addition, the overall accuracy (*ACC*), macro F1-averaging (*MF1*), and Cohen's kappa coefficient (κ) were also calculated to evaluate the performance of sleep scoring.

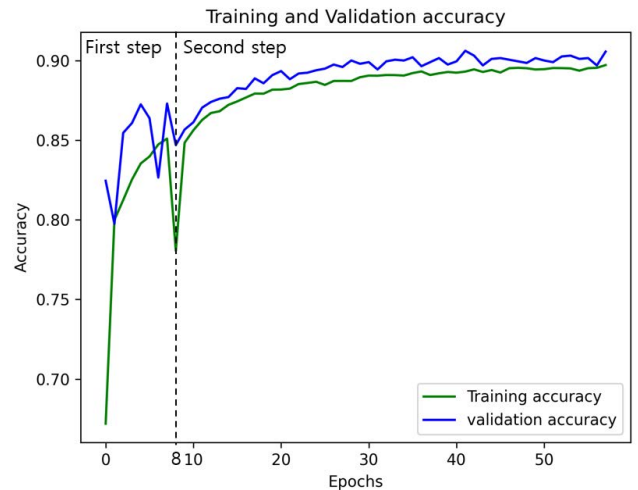$$ACC = \frac{\sum_{i=1}^{C} TP_i}{N} \quad (11)$$

FIGURE 7. Training and validation accuracy for random fold (i.e. first fold) in Sleep-EDF dataset. Learning epoch of First training step is 8 and second step is 50.

$$MF1 = \frac{\sum_{i=1}^{C} F1_i}{5} \quad (12)$$

$$\kappa = \frac{ACC - P_e}{1 - P_e},$$

$$P_e = \frac{\sum_{i=1}^{5}(TP_i + FN_i)(TP_i + FP_i)}{N^2} \quad (13)$$

where *C* denotes the number of classes and *N* the total number of test data. P_e denotes the hypothetical probability of the chance agreement [47].

III. RESULT

In this study, the proposed model was designed to classify the five sleep stages: Wake, N1, N2, N3, and REM stages. To evaluate the performance of the model, experiments were conducted using Fpz-Cz, Pz-Oz, and F3-M2 channel EEG signals from the Sleep-EDF datasets and WSC dataset, and several performance metrics were utilized for the evaluation.

A. CLASSIFICATION PERFORMANCE

Fig 7. illustrates the accuracy graph of the proposed model during the two-step training. The proposed model yields a stable convergence of the training and validation accuracy. Therefore, it demonstrates that the model is robust against overfitting.

There is an imbalance problem in the number of sleep stage labels. The epoch number of the N1 and N3 stages are smaller than those of the other stages in the dataset, as shown in Table 1. Thus, the F1 score and MF1 are important metrics to measure the performance of the imbalanced dataset as well as the overall accuracy. We verified the generalized performance of the model through a total of 5 experiments with three datasets using cross-validation. Table 2 shows the confusion matrices and performances of the proposed model using Fpz-Cz, Pz-Oz, and F3-M2 channel signals from the Sleep-EDF dataset and WSC dataset. The left and right parts in Table 2 show the confusion matrix and the performance
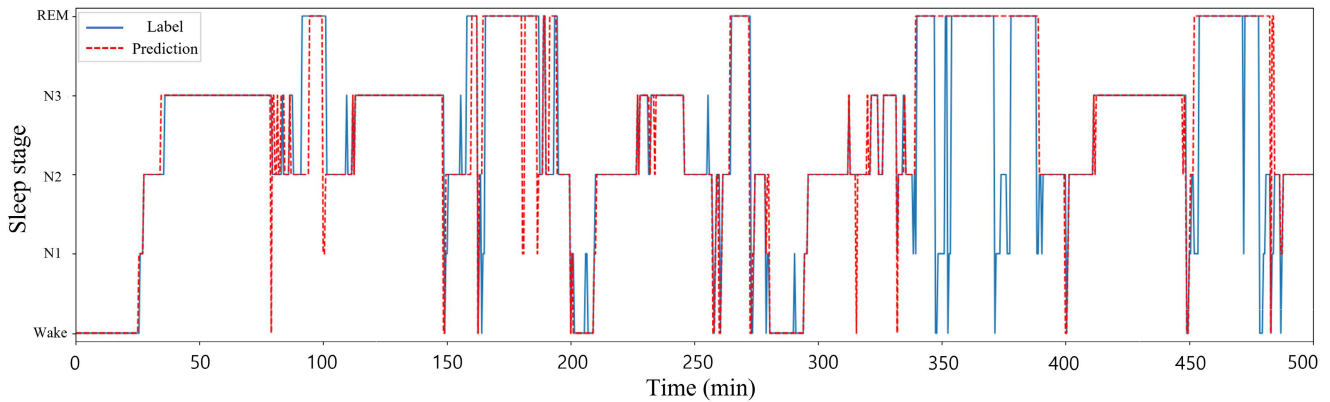
FIGURE 8. An example of sleep scoring by a sleep expert and our model using 'SC4001E0' participant's first night's record from the Sleep-EDF dataset. The blue line is the ground truth of the dataset, and the red dotted line is the sleep stage predicted by the proposed model.

TABLE 2. Confusion matrix and performance using each EEG channel from three datasets.

(a) Fpz-Cz channel signals of the Sleep-EDF-2013 dataset

	Predict					Per-class Performance(%)		
	Wake	N1	N2	N3	REM	PRE	RE	F1
Wake	**7380**	460	123	21	173	87.82	90.47	89.13
N1	407	**1299**	603	11	484	55.21	46.33	50.38
N2	374	376	**15786**	612	651	89.86	88.69	89.27
N3	31	3	511	**5158**	0	88.87	90.44	89.65
REM	212	215	545	2	**6743**	83.75	87.38	85.53
	ACC = 86.22%, MF1 = 80.79%, $\kappa = 0.81$							

(b) Pz-Oz channel signals of the Sleep-EDF-2013 dataset

	Predict					Per-class Performance(%)		
	Wake	N1	N2	N3	REM	PRE	RE	F1
Wake	**7342**	371	131	12	301	85.57	90.01	87.73
N1	664	**1017**	625	12	486	50.97	36.26	42.38
N2	304	367	**15647**	845	636	86.73	87.91	87.32
N3	28	1	966	**4706**	2	84.37	82.52	83.43
REM	242	239	671	3	**6562**	82.16	85.03	83.57
	ACC = 83.63% MF1 = 76.89% $\kappa = 0.77$							

(c) Fpz-Cz channel signals of the Sleep-EDF-2018 dataset

	Predict					Per-class Performance(%)		
	Wake	N1	N2	N3	REM	PRE	RE	F1
Wake	**61262**	3211	517	33	558	91.60	93.41	92.50
N1	3966	**9043**	6251	108	2011	49.31	42.30	45.53
N2	782	4191	**59215**	2049	1899	83.32	86.91	85.08
N3	40	24	2400	**10379**	14	81.90	80.73	81.31
REM	827	1871	2685	103	**20061**	81.74	78.26	80.10
	ACC = 82.67%, MF1 = 76.90%, $\kappa = 0.76$							

(d) Pz-Oz channel signals of the Sleep-EDF-2018 dataset

	Predict					Per-class Performance(%)		
	Wake	N1	N2	N3	REM	PRE	RE	F1
Wake	**61173**	3130	441	35	971	88.56	93.04	90.74
N1	4696	**8368**	6210	79	2116	51.08	38.98	44.22
N2	1734	3481	**57669**	2704	3045	81.05	86.80	82.51
N3	71	42	4077	**8545**	256	75.16	65.78	70.16
REM	1403	1360	2757	6	**20241**	76.01	78.55	77.26
	ACC = 80.16% MF1 = 72.98% $\kappa = 0.72$							

(e) F3-M2 channel signals of the WSC dataset

	Predict					Per-class Performance(%)		
	Wake	N1	N2	N3	REM	PRE	RE	F1
Wake	**19110**	890	782	26	202	92.31	91.95	92.13
N1	182	**2897**	2140	5	1326	48.95	43.02	45.79
N2	1556	917	**49537**	1193	1013	83.21	86.80	84.97
N3	46	6	2281	**5145**	29	79.58	81.07	80.32
REM	425	578	1220	2	**10842**	80.92	78.90	79.90
	ACC = 84.20%, MF1 = 76.62%, $\kappa = 0.75$							

metrics (PRE, RE, F1) of each sleep stage, and the overall performance of the model is described using *ACC*, *MF1* and κ across all sleep stages. In Table 2, the model trained with the Fpz-Oz channel performed better than that with the Pz-Oz channel in Sleep-EDF dataset. This result demonstrates that the Fpz-Cz channel has more distinct features than the Pz-Oz channel for sleep scoring. Due to the data imbalance issue among the classes in all the datasets, the overall accuracy is higher than 80%, whereas MF1 is relatively lower than 80%. In particular, the F1 score of the N1 stage is lower than those of the other stages owing to the fewer number of epochs. In contrast, the Wake and N2 stages with relatively a number of epochs yielded high F1 scores. As a result, the performance of sleep scoring using Fpz-Cz signals of the 2013 Sleep-EDF dataset and Pz-Oz signals of the 2018 Sleep-EDF dataset showed 80.79% and 72.98% MF1, i.e., the highest and lowest performance among the five experiments, respectively.

Additionally, the proposed model was tested for the separation between two class sleep stages (Wake, sleep), and among the three (Wake, NREM, REM), and four class sleep stages (Wake, Light Sleep, Deep-Sleep, REM), which can be seen in Table 3. The fewer classes were classified, the higher the performance of the proposed model was shown.

Fig 8. displays an example of automatic sleep scoring algorithms scored by an expert and the proposed model. Notably, most stages are well predicted except the N1 stage, which might be due to lower occurrence than the others.

B. BENCHMARKING WITH OTHER STATE-OF-THE-ART ALGORITHMS

In this section, we compare the sleep scoring performances of the proposed model with those of other state-of-the-art models. All studies used the same dataset and EEG channel and the same cross-validation method as ours. Table4 summarizes the benchmarking results of the per-class F1 score and overall performance metrics. Our proposed model using NA-BEMD has the highest overall accuracy and Cohen's kappa coefficient metrics for the three channel EEG dataset. Additionally, the MF1 score is the highest except the Pz-Oz channel EEG of the 2013 Sleep-EDF dataset, where our model has a higher F1 score in the N2 and REM stages than the other state-of-the-art algorithms and has a

TABLE 3. Performance of model according to the types of sleep stages to be classified.

(a) 4-class (Wake, Light Sleep, Deep-Sleep, REM)

Dataset	Channel	Overall Performance		
		ACC(%)	MF1(%)	k
Sleep-EDF 2013	Fpz-Cz / Pz-Oz	88.88 / 87.92	79.80 / 76.94	0.82 / 0.80
Sleep-EDF 2018	Fpz-Cz / Pz-Oz	84.97 / 83.64	77.00 / 75.29	0.78 / 75.76
WSC	F3-M2	88.91	78.72	0.81

(b) 3-class (Wake, NREM, REM)

Dataset	Channel	Performance		
		ACC(%)	MF1(%)	k
Sleep-EDF -2013	Fpz-Cz / Pz-Oz	91.23 / 90.30	89.43 / 88.16	0.84 / 0.82
Sleep-EDF -2018	Fpz-Cz / Pz-Oz	90.43 / 88.68	88.07 / 86.09	0.84 / 0.81
WSC	F3-M2	91.91	88.85	0.84

(c) 2-class (Wake, Sleep)

Dataset	Channel	Performance		
		ACC(%)	MF1(%)	k
Sleep-EDF -2013	Fpz-Cz / Pz-Oz	98.15 / 98.06	93.27 / 92.40	0.94 / 0.94
Sleep-EDF -2018	Fpz-Cz / Pz-Oz	97.77 / 97.65	94.38 / 93.06	0.95 / 0.95
WSC	F3-M2	98.14	93.91	0.94

comparable or higher F1 score in the Wake, N1, and N3 stages than the other models. As a result of the benchmark test using an FIR bandpass filter instead of NA-BEMD, up to 1% lower performance was observed than the model using NA-BEMD. Therefore, the proposed model proved that NA-BEMD is an efficient EEG decomposition method to classify the sleep stages rather than the Fourier-based frequency decomposition approach.

C. INTERPRETABLE ATTENTION WEIGHTS OF THE DECOMPOSED IMFs

The higher the attention weight given to the IMF, the more important the frequency component of the IMF is to classify the sleep stages. Each IMF has a certain frequency range based on the dyadic filter bank property of NA-BEMD (see Fig. 3.), and the spectrum includes a unique EEG frequency spectrum. IMF2 contains the beta (13-30Hz) and alpha band (8-13Hz) frequency components, IMF3 alpha, theta bands (4-8Hz), IMF4 little alpha, and theta and delta bands (0.5-4Hz), and IMF5 and IMF6 mostly include delta band frequency components. Fig 9. displays the mean and standard deviation of the attention weights given to the IMFs for each sleep stage. IMF2, which has the widest spectrum including high-frequency components, had the highest mean of attention weight in all stages. Notably, the highest mean of attention weight was assigned to the Wake stage among the mean of attention weights of IMF2. Also, the mean of attention weights of IMF3 and IMF4, which contains theta band spectrum, decreased from Wake stage to N3 stage In particular, IMF4, which contains a strong theta band spectrum, had higher mean of attention weights for N1 stage and N2 stage than that of IMF3. Additionally, the mean of attention weights of IMF5 for the N3 stage epochs, which contains a strong delta band spectrum, were higher
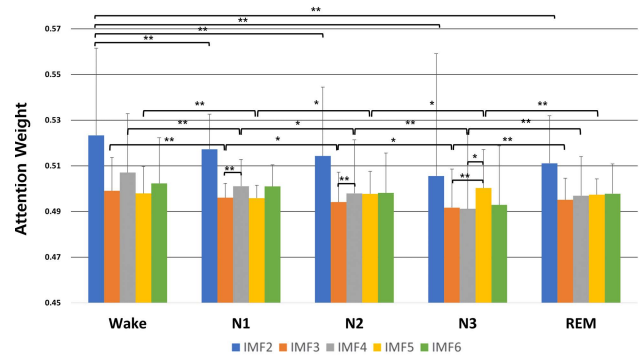

FIGURE 9. Attention weight of the IMFs for each sleep stage. The mean and standard deviation were calculated using the attention weights extracted from the test set of the first fold during the cross-validation of the Sleep-EDF-2013 dataset. One asterisk (*) indicates p-value smaller than 0.05 ($p < 0.05$). Two asterisks (**) indicate ($p < 0.01$).

than other stages. One-tailed t-test was conducted to confirm whether these differences were significant. Interestingly, these attention weights to the different IMFs are consistent with the findings of the previous studies [28]. In addition, it was confirmed that each sleep stage does not depend on only one IMF, but multiple IMFs contribute to a sleep stage.

IV. DISCUSSION

Compared to the existing sleep scoring studies, the main contributions of the proposed method are the applications of NA-BEMD to extract the EEG frequency components and the attention mechanism to emphasize the corresponding frequency components to each sleep stage. Additionally, a novel deep neural network consisting of 1D-CNN and Bi-LSTM, and interpretability with the attention mechanism are the main contributions to the automatic sleep scoring research. These approaches are robust and reliable for the sleep scoring using only a single channel EEG signal.

Since the sleep experts consider the slow and fast scales in EEG signals as well as their amplitudes when classifying the sleep stages, the automatic sleep scoring algorithm also needs to decompose an EEG signal into several frequency components and consider the significance of the components corresponding to the sleep stages. For the frequency decomposition, the traditional Fourier-based method could be utilized using its sine and cosine basis functions. However, due to the limiation of its basis functions, it would not handle the inherent non-linearity and non-stationarity of EEG signals. To solve this issue, the data-driven decomposition algorithm, EMD, has been applied. Furthermore, in order to utilize the advantage of the dyadic filter bank property, NA-BEMD is eventually utilized. The benchmark test of the sleep scoring algorithms using NA-BEMD and FIR filters demonstrates NA-BEMD outperforms the other.

The attention mechanism for the decomposed frequency components as inputs to the proposed model yield high weights to the components with large information about a sleep stage, which could provide interpretability regarding the proposed model and enhance the performance. For instance,

TABLE 4. Bench mark results compared with other studies based on single channel EEG.

Dataset	Channel	Methods	Overall Metrics			Per-class F1 score(%)				
			ACC(%)	MF1(%)	k	Wake	N1	N2	N3	REM
Sleep-EDF -2013	Fpz-Cz	Back et al. [48]	84.0	77.7	0.78	87.9	44.7	88.0	85.7	82.1
		Supratak et al. [20]	82.0	76.9	0.76	84.7	46.6	85.9	84.8	82.4
		Zhu et al. [19]	82.8	77.8	-	**90.3**	47.1	86.0	82.1	83.2
		Mousavi et al. [24]	84.3	79.7	0.79	89.2	**52.2**	86.8	85.1	85.0
		Neng et al. [25]	84.3	79.8	0.78	89.0	51.7	87.3	88.2	82.9
		Eldele et al. [26]	84.4	78.1	0.79	89.7	42.6	88.8	**90.2**	79.0
		proposed (BF)	85.7	78.0	0.80	88.9	48.1	88.9	89.0	85.0
		proposed (NA-BEMD)	**86.2**	**80.8**	**0.81**	89.1	50.4	**89.3**	89.7	**85.5**
	Pz-Oz	Supratak et al. [20]	79.8	73.1	0.72	88.1	37	82.7	77.3	80.3
		Mousavi et al. [24]	82.8	**77.0**	**0.77**	**90.3**	**44.6**	85.7	81.6	82.9
		Neng et al. [25]	80.3	74.6	0.73	86.0	41.5	84.9	81.0	79.6
		proposed (BF)	83.0	76.1	0.76	86.8	40.9	87.0	83.4	82.5
		proposed (NA-BEMD)	**83.6**	**77.0**	**0.77**	87.7	42.4	**87.4**	83.5	83.6
Sleep-EDF -2018	Fpz-Cz	Mousavi et al. [24]	80.0	73.6	0.73	91.7	44.1	82.5	73.6	76.1
		Eldele et al. [26]	81.3	75.1	0.74	92.0	42.0	85.0	**82.1**	74.2
		proposed (BF)	81.6	76.0	0.75	92.0	43.9	84.8	80.3	79.4
		proposed (NA-BEMD)	**82.7**	**76.9**	**0.76**	**92.5**	45.5	**85.1**	81.3	**80.1**
	Pz-Oz	Mousavi et al. [24]	77.6	70.0	0.69	90.3	42.2	79.7	**94.8**	72.2
		proposed (BF)	80.0	72.8	**0.72**	90.4	**44.3**	**82.6**	70.2	76.7
		proposed (NA-BEMD)	**80.2**	**73.0**	**0.72**	**90.7**	44.2	82.5	70.2	**77.3**
WSC	F3-M2	Mousavi et al. [24]	74.0	65.8	61.8	82.4	38.2	81.2	64.0	63.1
		proposed (BF)	84.0	74.8	**0.75**	86.6	41.9	**90.0**	74.4	**81.4**
		proposed (NA-BEMD)	**84.2**	**74.9**	**0.75**	**92.1**	**45.8**	85.0	80.3	79.9

since IMF2 considerably varies among the sleep stages, it could be crucial to classify the sleep stages. Thus, the attention weights of IMF2 are high in all sleep stages, which contain high-frequency components such as alpha or beta bands, contributing the most to the classification of all sleep stages. For the same reason, we found that the IMF5, which contains low-frequency components such as the delta band, contributes the most to the N3 stage classification, and other IMFs contribute to the wake stage. Therefore, this method not only improve the model performance, but also makes it reliable with the explainability of the designed deep neural networks.

Another main novelty of the proposed model is the combination of the 1D-CNN and Bi-LSTM to extract the features in each epoch and the continuous epochs. The relations among the EEG epochs are crucial to define the sleep stages due to the nature of the sleep cycles [49]. Therefore, the automatic sleep scoring model is designed to consider this in the hidden state during the learning process. To improve the performance of the model, a deeper 1D-CNN layer with the residual connection is built, which also prevents the overfitting of 1D-CNN through two-step training. As a result, the proposed model outperformed the other state-of-the-art methods in the five experiments the same channel of EEG signal from the same dataset.

In the experiment, inter-participant cross validation was conducted using the data of three EEG channels (Fpz-Cz, Pz-Oz, and F3-M2) from the three public datasets. In particular, the F3-M2 channel is useful for the implementation of sleep

monitoring wearable devices. It can be recorded from two areas of the head, that is, behind the ear and in the front of the head, which is hairless and more comfortable to be attached than the other channels. In addition, because the WSC dataset has data from participants of various races —i.e., Asian, Black, Hispanic, Native American, and White— the experiments using the WSC dataset demonstrated that our proposed model could be generalized for various ethnicities.

Although our model showed the best results with the most performance metrics, there were a limitations. As the number of N1 epochs is insufficient, the prediction probability of the N1 stage was still low, and the proposed model could not learn features of sleep spindle or k-complex separately. Therefore, for future works, we need to improve the performance of the N1 scoring using a rule-based method which reflects the N1 and N2 stage transition probability [50] or oversamples augmenting N1 epochs [51]. Also, the proposed model could be improved in performance by detecting the sleep spindle or k-complex and extracting features in future works.

V. CONCLUSION

In this study, we propose a robust model for an automatic sleep scoring method using an single channel EEG. In contrast to other state-of-the-art models, the proposed model uses the noise-assisted bivariate EMD for the extraction of the intrinsic modes of the EEG signals and applies a residual connection and various filter sizes for the CNN architecture to design a deeper and more robust model to

improve the representative feature extraction performance. Additionally, to learn the relationship between consecutive epochs, 10 epochs are used as one sentence and are learned using Bi-LSTM. Furthermore, the model was interpretable using the attention mechanism. Based on these findings, we prove that the proposed model shows the best performance for three channel EEG signals from the three public datasets compared with the other state-of-the-art algorithms. Therefore, we expect that our robust model can be practically applied to professional decision support systems and wearable sleep monitoring devices.

REFERENCES

[1] S. Chokroverty, "Overview of sleep & sleep disorders," *Indian J. Med. Res.*, vol. 131, no. 2, pp. 126–140, 2010.

[2] M. Engstrøm, E. Rugland, and M. S. Heier, "Polysomnography (PSG) for studying sleep disorders," *Tidsskrift Norske laegeforening, Tidsskrift Praktisk Medicin*, vol. 133, no. 1, pp. 58–62, 2013.

[3] T. Hori, Y. Sugita, E. Koga, S. Shirakawa, K. Inoue, S. Uchida, H. Kuwahara, M. Kousaka, T. Kobayashi, Y. Tsuji, M. Terashima, K. Fukuda, and N. Fukuda, "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 Clin. Neurosci.*, vol. 55, no. 3, pp. 305–310, Jun. 2001.

[4] R. B. Berry, R. Brooks, C. Gamaldo, S. M. Harding, R. M. Lloyd, S. F. Quan, M. T. Troester, and B. V. Vaughn, "AASM scoring manual updates for 2017," *J. Clin. Sleep Med.*, vol. 13, no. 5, pp. 665–666, May 2017.

[5] H. Danker-Hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, and G. Dorffner, "Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard," *J. Sleep Res.*, vol. 18, no. 1, pp. 74–84, Mar. 2009.

[6] A. Malhotra, M. Younes, S. T. Kuna, R. Benca, C. A. Kushida, J. Walsh, A. Hanlon, B. Staley, A. I. Pack, and G. W. Pien, "Performance of an automated polysomnography scoring system versus computer-assisted manual scoring," *Sleep*, vol. 36, no. 4, pp. 573–582, Apr. 2013.

[7] I. Hussain and S. J. Park, "Big-ECG: Cardiographic predictive cyber-physical system for stroke management," *IEEE Access*, vol. 9, pp. 123146–123164, 2021.

[8] I. Hussain and S. J. Park, "HealthSOS: Real-time health monitoring system for stroke prognostics," *IEEE Access*, vol. 8, pp. 213574–213586, 2020.

[9] K. K. Krishnan and K. Soman, "CNN based classification of motor imaginary using variational mode decomposed EEG-spectrum image," *Biomed. Eng. Lett.*, vol. 11, no. 3, pp. 235–247, 2021.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25, Dec. 2012, pp. 1097–1105.

[11] V. Y. Senyurek, M. H. Imtiaz, P. Belsare, S. Tiffany, and E. Sazonov, "A CNN-LSTM neural network for recognition of puffing in smoking episodes using wearable sensors," *Biomed. Eng. Lett.*, vol. 10, no. 2, pp. 195–203, May 2020.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] E. Oropesa, H. L. Cycon, and M. Jobert, "Sleep stage classification using wavelet transform and neural network," Int. Comput. Sci. Inst., Berkeley, CA, USA, Tech. Rep., Mar. 1999.

[14] K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, H. Dickhaus, and L. Fraiwan, "Classification of sleep stages using multi-wavelet time frequency entropy and LDA," *Methods Inf. Med.*, vol. 49, no. 3, pp. 230–237, 2010.

[15] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012.

[16] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2959–2978, Oct. 2017.

[17] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[18] O. Yildirim, U. Baloglu, and U. Acharya, "A deep learning model for automated sleep stages classification using PSG signals," *Int. J. Environ. Res. Public Health*, vol. 16, no. 4, p. 599, Feb. 2019.

[19] T. Zhu, W. Luo, and F. Yu, "Convolution- and attention-based neural network for automated sleep stage classification," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, p. 4152, Jun. 2020.

[20] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.

[21] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "DNN filter bank improves 1-max pooling CNN for single-channel EEG automatic sleep stage classification," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 453–456.

[22] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Comput. Biol. Med.*, vol. 106, pp. 71–81, Mar. 2019.

[23] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 400–410, Mar. 2019.

[24] S. Mousavi, F. Afghah, and U. R. Acharya, "SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0216456.

[25] W. Neng, J. Lu, and L. Xu, "CCRRSleepNet: A hybrid relational inductive biases network for automatic sleep stage classification on raw single-channel EEG," *Brain Sci.*, vol. 11, no. 4, p. 456, Apr. 2021.

[26] E. Eldele, Z. Chen, C. Liu, M. Wu, C.-K. Kwoh, X. Li, and C. Guan, "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 809–818, 2021.

[27] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Process. Mag.*, vol. 9, no. 2, pp. 21–67, Apr. 1992.

[28] A. Brancaccio, D. Tabarelli, M. Bigica, and D. Baldauf, "Cortical source localization of sleep-stage specific oscillatory activity," *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, Dec. 2020.

[29] A. R. Hassan, M. Imamul, and M. Bhuiyan, "Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting," *Comput. Methods Programs Biomed.*, vol. 140, pp. 201–210, Mar. 2017.

[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[31] A. L. Goldberger, L. A. N. 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, Jun. 2000.

[32] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline, "The national sleep research resource: Towards a sleep data commons," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1351–1358, Oct. 2018.

[33] L. Fiorillo, A. Puiatti, M. Papandrea, P.-L. Ratti, P. Favaro, C. Roth, P. Bargiotas, C. L. Bassetti, and F. D. Faraci, "Automated sleep scoring: A review of the latest approaches," *Sleep Med. Rev.*, vol. 48, Dec. 2019, Art. no. 101204.

[34] C. S. Herrmann, D. Strüber, R. F. Helfrich, and A. K. Engel, "EEG oscillations: From correlation to causality," *Int. J. Psychophysiol.*, vol. 103, pp. 12–21, May 2016.

[35] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, pp. 903–995, Mar. 1971.

[36] Y. Kim, J. Ryu, K. K. Kim, C. C. Took, D. P. Mandic, and C. Park, "Motor imagery classification using mu and beta rhythms of EEG with strong uncorrelating transform based complex common spatial patterns," *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–13, Oct. 2016.

[37] P. Vardhini, N. Punitha, and S. Ramakrishnan, "Empirical mode decomposition based detrended fluctuation analysis of uterine electromyography signals in term conditions," *IEIE Trans. Smart Process. Comput.*, vol. 9, no. 2, pp. 127–134, Apr. 2020.

[38] G. Rilling, P. Flandrin, P. Gonçalves, and J. M. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 936–939, Dec. 2007.

[39] N. U. Rehman, C. Park, N. E. Huang, and D. P. Mandic, "EMD via MEMD: Multivariate noise-aided computation of standard EMD," *Adv. Adapt. Data Anal.*, vol. 5, no. 2, 2013, Art. no. 1350007.

[40] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Müller, and F. Petitjean, "InceptionTime: Finding AlexNet for time series classification," *Data Mining Knowl. Discovery*, vol. 34, no. 6, pp. 1936–1962, Sep. 2019.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[42] M. Laavanya and V. Vijayaraghavan, "Residual learning of transfer-learned AlexNet for image denoising," *IEIE Trans. Smart Process. Comput.*, vol. 9, no. 2, pp. 135–141, Apr. 2020.

[43] A. A. Borbély, F. Baumann, D. Brandeis, I. Strauch, and D. Lehmann, "Sleep deprivation: Effect on sleep stages and EEG power density in man," *Electroencephalogr. Clin. Neurophysiol.*, vol. 51, no. 5, pp. 483–493, May 1981.

[44] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[45] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc. B, Methodol.*, vol. 36, no. 2, pp. 111–133, 1974.

[46] K. A. I. Aboalayon, H. T. Ocbagabir, and M. Faezipour, "Efficient sleep stage classification based on EEG signals," in *Proc. IEEE Long Island Syst., Appl. Technol. (LISAT) Conf.*, May 2014, pp. 1–6.

[47] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[48] H. Seo, S. Back, S. Lee, D. Park, T. Kim, and K. Lee, "Intra- and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG," 2019, *arXiv:1902.06562*.

[49] I. Feinberg and T. C. Floyd, "Systematic trends across the night in human sleep cycles," *Psychophysiology*, vol. 16, no. 3, pp. 283–291, May 1979.

[50] H. Yu, O.-S. Shin, S. Kim, and C. Park, "Effect of an inflatable air mattress with variable rigidity on sleep quality," *Sensors*, vol. 20, no. 18, p. 5317, Sep. 2020.

[51] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 28, pp. 321–357, Jun. 2006.

JAEWOO BAEK received the B.S. degree in computer engineering from Kwangwoon University, Seoul, South Korea, where he is currently pursuing the master's degree in computer engineering. His research interests include biological signal processing, machine learning, deep learning, and reinforcement learning.

CHOONGSEOP LEE received the B.S. degree in computer science and engineering from Kwangwoon University, Seoul, South Korea. His research interests include machine learning and computational neuroscience.

HYUNSOO YU received the B.S. degree in robotics engineering from Kwangwoon University, Seoul, South Korea, where he is currently pursuing the master's degree in computer engineering. His research interests include in experimental setting, signal processing, machine learning, and artificial intelligence.

SUWHAN BAEK is currently pursuing the master's degree with Kwangwoon University, Seoul, Republic of Korea. He is also researching the Green AI field with the University of Bristol, U.K., as a Visiting Researcher. His research interests include overall medical AI, reinforcement learning, and safe AI.

SEOKMIN LEE received the B.S. degree in computer information engineering from Kwangwoon University, Seoul, South Korea, in 2022. He is currently pursuing the master's degree in convergence security with Korea University. His research interests include the security of CPU microarchitecture, cloud environments, and artificial intelligence.

SEUNGMIN LEE received the B.S. degree in electronics engineering from Sogang University, Seoul, South Korea, in 2005, and the M.S. and Ph.D. degrees in interdisciplinary program for bioengineering from Seoul National University, Seoul, in 2007 and 2012, respectively. The thesis of his Ph.D. degree was focused on the nerve recording using implantable devices and analyzing physiological signal. He is currently an Assistant Professor with Kookmin University, Seoul. His research interests include biosignal processing, active microelectrodes, new electronic materials, and electronic devices for biomedical engineering.

CHEOLSOO PARK (Member, IEEE) received the B.Eng. degree in electrical engineering from Sogang University, Seoul, the M.Sc. degree from the Biomedical Engineering Department, Seoul National University, South Korea, and the Ph.D. degree in adaptive nonlinear signal processing from the Imperial College London, U.K., in 2012. He is currently an Associate Professor with the Computer Engineering Department, Kwangoon University, Seoul, South Korea. He worked as a Postdoctoral Researcher with the Bioengineering Department, University of California at San Diego, San Diego, USA. His research interests include the areas of machine learning and adaptive and statistical signal processing, with applications in brain–computer interfaces, computational neuroscience, and wearable technology.

● ● ●